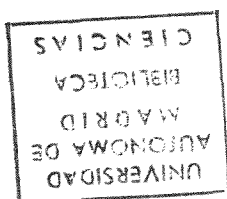


717954

5/1380

SECUENCIA DE NUCLEOTIDOS DE LOS EXTREMOS

DEL GENOMA DEL VIRUS DE LA PESTE PORCINA AFRICANA



Py F e  
10536

Tesis presentada por  
D. Víctor Calvo López  
para optar al grado de  
doctor en Ciencias.

Madrid, Junio 1986

SECUENCIA DE NUCLEOTIDOS DE LOS EXTREMOS  
DEL GENOMA DEL VIRUS DE LA PESTE PORCINA AFRICANA

VICTOR CALVO LÓPEZ

## ABREVIATURAS.

Ac: anticuerpo.

AcNa: acetato sódico.

Ag: antígeno.

ATP: adenosina-5'-trifosfato.

Bisacrilamida: N,N'-metilenbisacrilamida.

BPB: azul de bromofenol.

BSA: seralbúmina bovina.

Ci: Curio.

CTAB: bromuro de cetiltrimetilamonio.

ddATP: didesoxiadenosina-5'-trifosfato.

DMS: sulfato de dimetilo.

DMSO: dimetilsulfóxido.

DNA: ácido desoxirribonucleico.

DNAsa: desoxirribonucleasa.

DNAts: DNA de esperma de salmón.

dNTP: desoxinucleósido-5'-trifosfato.

DTT: D-L-ditiotreitol.

EDTA: ácido etilendiaminotetraacético.

EtBr: bromuro de etidio.

EtOH: etanol.

HZ: hidracina.

IPTG: isopropil-β-D-tiogalactpiranósido.

K: kilodaltons.

kb: kilobases.

kpb: kilopares de bases.

2-ME: 2-mercaptoetanol.

NTP: nucleósido-5'-trifosfato.

pI: punto isoelectrico.

Pm: peso molecular.

PB: poxvirus bovino.

RNAse: ribonucleasa.

rpm: revoluciones por minuto.

PVP: polivinilpirrolidona.

Sarcosil: N-lauroilsarcosinato sódico.

SCM: sitio de clonaje múltiple.

SDS: dodecilsulfato sódico.

ta: temperatura ambiente.

TEMED: N,N,N',N'-tetrametiletildiamina.

Tris: trimetilhidroxiaminometano.

tRNA: RNA de transferencia.

u: unidad.

ug: microgramo.

ul: microlitro.

VPPA: virus de la peste porcine africana.

VV: virus de la vacuna.

XC: xilencianol.

X-gal: 5-bromo-4-cloro-3-indolil- $\beta$ -D-galactopiranosido.



## Endonucleasas de restricción.

A: AccI.

B: BamHI; Bc: BclI; Bg: BglII; Bs: BssHII.

D: DraI.

H: HindIII; Hc: HincII.

K: KpnI.

M: SmaI.

N: NcoI; Ns: NsiI.

P: PvuI; Ps: PstI; Pv: PvuII.

R: EcoRI.

S: Sac I; Sc: ScaI; St:StuI.

T: TaqI.

V: EcoRV.

X: XbaI.

## INTRODUCCION

## 1. PESTE PORCINA AFRICANA.

La peste porcina africana (PPA) es una enfermedad altamente contagiosa que afecta al cerdo doméstico y especies relacionadas, produciendo una elevada mortalidad, próxima al 100%, en sus formas virulentas.

### 1.1. Etiología e historia.

El agente etiológico es un virus (1,2) que puede aislarse de sangre, órganos internos y de todas las secreciones y excreciones del animal infectado (3). Este virus constituye una amenaza para la cabaña porcina mundial debido a la inexistencia de vacuna para el mismo, a su capacidad para multiplicarse en parásitos del cerdo (garrapatas) y cambiar rápidamente, y al hecho de que diferentes formas del virus pueden producir enfermedades con diferentes sintomatologías o carentes totalmente de síntomas clínicos.

La naturaleza viral de la enfermedad y las diferencias inmunológicas respecto a la peste porcina clásica fueron descritas por Montgomery en el año 1921 (2) al resumir sus trabajos efectuados en Kenya.

La enfermedad apareció por primera vez fuera del continente africano en el año 1957 en las proximidades de Lisboa (4). Aunque este primer brote fue eliminado aparentemente, la enfermedad reapareció en Portugal en 1960, pasó a España

y desde entonces es enzoótica en la Península Ibérica (4). En los años subsiguientes (1964-1980) se registraron brotes epizooticos en Francia, Italia y Malta. En América los brotes aparecidos se circunscriben geográficamente al área del Caribe (Cuba, República Dominicana, Haití) y a Brasil, y se produjeron más recientemente (a partir de 1971).

### 1.2. Patología e inmunología.

La enfermedad se puede presentar actualmente en 4 formas: hiperaguda, aguda, subaguda y crónica (4). Estas diversas formas clínicas hacen arbitraria cualquier determinación de un cuadro clínico específico y universal para la PPA. La forma clínica en que se presenta la enfermedad viene determinada fundamentalmente por el virus, pero puede ser influida por el estado fisiológico general del animal (5).

Mientras que la muerte puede ser la única manifestación de los casos hiperagudos, en los casos agudos la primera señal es fiebre, seguida de anorexia, incoordinación muscular y tambaleo como síntomas más comunes, aunque no únicos. Todos estos síntomas acentuados se pueden presentar en la forma hiperaguda. El análisis de los animales muertos revela hemorragias generalizadas en los nódulos linfáticos, riñón, corazón estómago y colon (6). Este cuadro clínico suele hacerse corresponder a la forma subaguda cuando los síntomas son menos intensos. Al igual que en el cuadro agudo, la muerte parece estar producida por daños en el sistema vascu-

y desde entonces es enzoótica en la Península Ibérica (4). En los años subsiguientes (1964-1980) se registraron brotes epizooticos en Francia, Italia y Malta. En América los brotes aparecidos se circunscriben geográficamente al área del Caribe (Cuba, República Dominicana, Haití) y a Brasil, y se produjeron más recientemente (a partir de 1971).

### 1.2. Patología e inmunología.

La enfermedad se puede presentar actualmente en 4 formas: hiperaguda, aguda, subaguda y crónica (4). Estas diversas formas clínicas hacen arbitraria cualquier determinación de un cuadro clínico específico y universal para la PPA. La forma clínica en que se presenta la enfermedad viene determinada fundamentalmente por el virus, pero puede ser influida por el estado fisiológico general del animal (5).

Mientras que la muerte puede ser la única manifestación de los casos hiperagudos, en los casos agudos la primera señal es fiebre, seguida de anorexia, incoordinación muscular y tambaleo como síntomas más comunes, aunque no únicos. Todos estos síntomas acentuados se pueden presentar en la forma hiperaguda. El análisis de los animales muertos revela hemorragias generalizadas en los nódulos linfáticos, riñón, corazón estómago y colon (6). Este cuadro clínico suele hacerse corresponder a la forma subaguda cuando los síntomas son menos intensos. Al igual que en el cuadro agudo, la muerte parece estar producida por daños en el sistema vascu-

lar. Las hemorragias son más intensas en nódulos linfáticos y riñones.

En su forma crónica la PPA es difícil de reconocer, aparece en áreas enzoóticas, y los animales que la padecen son sobrevivientes de infecciones agudas (1,5), actuando como agentes transmisores de virus infectivos a lo largo de su vida.

La característica inmunológica más importante en la infección del VPPA es la falta de neutralización del virus por los anticuerpos producidos en respuesta a la infección, aunque estos son capaces de precipitarlo y dar lugar a complejos que fijan complemento (6). Animales inoculados en el laboratorio pueden producir altos niveles de anticuerpos fijadores de complemento (7-9), precipitantes (10-12) y neutralizantes contra otros virus (13), así como hipersensibilidad a otros antígenos (14). No obstante no se han encontrado en estos animales anticuerpos neutralizantes ni protectores contra el VPPA.

Para explicar este fenómeno existen diversas hipótesis, entre ellas la diversidad antigénica del virus utilizado en los ensayos de neutralización o la inducción de anticuerpos bloqueantes (15).

Otra posible explicación de la peculiar respuesta inmune frente al VPPA sería la existencia de competición antigénica, en la que un epítipo irrelevante pero dominante suprime la respuesta frente a un epítipo crítico. Esta competición podría ser intermolecular, o intramolecular, si los

epitopos en cuestión estuviesen localizados en distintas moléculas o en la misma, respectivamente.

Recientemente se ha descrito una neutralización del VPPA superior al 90% empleando antisueros de conejos hiperinmunizados y un anticuerpo monoclonal de raton dirigido contra una proteína presente tanto en células infectadas como en no infectadas (16). Sin embargo la neutralización del virus es incompleta pese a haberse obtenido anticuerpos monoclonales contra alrededor de 12 proteínas del VPPA (17).

## 2. LA PARTICULA VIRAL.

### 2.1. Morfología.

El virus de la PPA se ha clasificado como miembro de la familia Iridoviridae ,que incluye deoxivirus de insectos y de vertebrados, los cuales se caracterizan por una simetría aparentemente icosaédrica con un nucleoide rodeado por una unidad de membrana (18,19). Los iridovirus que infectan vertebrados poseen envueltas externas derivadas de la salida por gemación a través de la membrana plasmática (20-22).

El análisis por microscopía electrónica de los viriones del virus de la PPA preparados por tinción negativa, corte ultrafino, liofilización y sombreado muestran que la partícula viral está compuesta de varias estructuras concéntricas con forma icosaédrica (23). La región interna de las partículas virales es un nucleoide rodeado por una membrana cubierta por la cápsida. La cápsida tiene una dimensiones de

epitopos en cuestión estuviesen localizados en distintas moléculas o en la misma, respectivamente.

Recientemente se ha descrito una neutralización del VPPA superior al 90% empleando antisueros de conejos hiperinmunizados y un anticuerpo monoclonal de raton dirigido contra una proteína presente tanto en células infectadas como en no infectadas (16). Sin embargo la neutralización del virus es incompleta pese a haberse obtenido anticuerpos monoclonales contra alrededor de 12 proteínas del VPPA (17).

## 2. LA PARTICULA VIRAL.

### 2.1. Morfología.

El virus de la PPA se ha clasificado como miembro de la familia Iridoviridae ,que incluye deoxivirus de insectos y de vertebrados, los cuales se caracterizan por una simetría aparentemente icosaédrica con un nucleoide rodeado por una unidad de membrana (18,19). Los iridovirus que infectan vertebrados poseen envueltas externas derivadas de la salida por gemación a través de la membrana plasmática (20-22).

El análisis por microscopía electrónica de los viriones del virus de la PPA preparados por tinción negativa, corte ultrafino, liofilización y sombreado muestran que la partícula viral está compuesta de varias estructuras concéntricas con forma icosaédrica (23). La región interna de las partículas virales es un nucleoide rodeado por una membrana cubierta por la cápsida. La cápsida tiene una dimensiones de



172 a 191 nm, tamaño intermedio entre el virus de la linfocitosis (300 nm) y FV3 (130 nm), y está constituida por capsómeros ordenados en una trama hexagonal. Las micrografías de cápsidas sombreadas o tratadas por tinción negativa, filtradas por ordenador, revelan capsómeros con una estructura hexagonal y un orificio en el centro. La distancia intercapsómero varía de 7.4 a 8.1 nm. El número de triangulación de la cápsida se estimó en 189 a 217, indicativo de 1892 a 2172 capsómeros (23). El VPPA extracelular tiene una membrana externa análoga a la unidad de membrana citoplasmática. Esta unidad de membrana se puede eliminar por choque osmótico o detergentes, lo que sugiere que está formada por una envuelta lipídica. Este modelo presenta muchas analogías con el propuesto para FV3 (24).

## 2.2. Proteínas estructurales.

La caracterización bioquímica del VPPA se ha visto frenada tradicionalmente por la falta de un método adecuado de purificación de la partícula viral, por motivos similares a los que dificultan la purificación de los herpesvirus (25): la pérdida de infectividad de los viriones en condiciones utilizadas normalmente en purificaciones de virus, tales como elevada fuerza iónica (26); la contaminación del material de partida con vesículas membranosas de difícil eliminación (27,28), o la presencia en el medio extracelular de una mezcla de partículas incompletas, procedentes de lisis celular, y partículas completas que portan membrana externa adquirida

por gemación a través de la membrana citoplásmica de las células infectadas (29,30).

Además, las células porcinas susceptibles para aislados de campo del VPPA son monocitos y macrófagos (31), difíciles de obtener en grandes cantidades (32). Por esta razón todos los métodos de purificación descritos parten de virus adaptados a diversas líneas celulares establecidas (33-37).

Un avance importante en la purificación fué la introducción de Tween 80 y NaCl 1 M para disociar los viriones de los componentes celulares (34). Posteriormente el empleo de gradientes preformados de sacarosa permitió obtener un virus extracelular parcialmente purificado (35). Actualmente el método más efectivo (28) implica dos gradientes sucesivos de densidad en Percoll al 45% y el paso a través de columnas de Sephacryl S.1000 para eliminar las partículas de Percoll, todo lo cual permite alcanzar un preparado viral extracelular con una contaminación de vesículas relativa al número de partículas virales inferior al 5%.

El virus intracelular parcialmente purificado contiene 24 proteínas, una de las cuales es probablemente actina celular (38,39). Este número de proteínas es considerablemente mayor que el detectado por otros autores (34,36). Los pesos moleculares de dichas proteínas oscilan entre 10 y 243 kD. Analizando partículas virales extracelulares purificadas por Percoll se detectan 34 proteínas de pesos moleculares comprendidos entre 10 y 150 kD (28), de las cuales 20 son muy

por gemación a través de la membrana citoplásmica de las células infectadas (29,30).

Además, las células porcinas susceptibles para aislados de campo del VPPA son monocitos y macrófagos (31), difíciles de obtener en grandes cantidades (32). Por esta razón todos los métodos de purificación descritos parten de virus adaptados a diversas líneas celulares establecidas (33-37).

Un avance importante en la purificación fué la introducción de Tween 80 y NaCl 1 M para disociar los viriones de los componentes celulares (34). Posteriormente el empleo de gradientes preformados de sacarosa permitió obtener un virus extracelular parcialmente purificado (35). Actualmente el método más efectivo (28) implica dos gradientes sucesivos de densidad en Percoll al 45% y el paso a través de columnas de Sephacryl S.1000 para eliminar las partículas de Percoll, todo lo cual permite alcanzar un preparado viral extracelular con una contaminación de vesículas relativa al número de partículas virales inferior al 5%.

El virus intracelular parcialmente purificado contiene 24 proteínas, una de las cuales es probablemente actina celular (38,39). Este número de proteínas es considerablemente mayor que el detectado por otros autores (34,36). Los pesos moleculares de dichas proteínas oscilan entre 10 y 243 kD. Analizando partículas virales extracelulares purificadas por Percoll se detectan 34 proteínas de pesos moleculares comprendidos entre 10 y 150 kD (28), de las cuales 20 son muy

probablemente proteínas estructurales del virus, siendo de estas las mayoritarias p150, p72, p37, p35, p17, p12 y p10.

De las 34 proteínas detectadas en viriones purificados por Percoll, alrededor de 12 son probablemente componentes celulares, algunas de las cuales, si no todas, se encuentran incorporadas en el virión (28).

Se ha descrito la presencia en el virus intracelular de 3 glicoproteínas: gp 89, gp 56 y gp 51 (38,39). Sin embargo, usando virus purificado por Percoll no se ha detectado ninguna glicoproteína estructural en el virión (40).

La composición proteica de las diferentes capas del virión se ha estudiado tratando virus intracelular con NP40 y 2-mercaptoetanol (39) en ausencia o presencia de NaCl 0.5 M. En el primer caso se obtienen partículas subvirales (Core I) que no son infectivas y han perdido algunas proteínas. Tras la adición de NaCl, el Core I pierde su proteína mayoritaria, la p72, y da lugar a una subpartícula más pequeña (Core II).

### 2.3. Actividades enzimáticas.

El virus de la PPA contiene una RNA-polimerasa dependiente del DNA (37), similar a la descrita en poxvirus (41,42), que sintetiza RNAs poliadenilados y con "CAP". Este hecho implica que en la partícula viral deben encontrarse los enzimas responsables de estas modificaciones: poli (A) polimerasa, guanililtransferasa y metiltransferasa (43).

Otras actividades presentes son una proteinquinasa (35), una o posiblemente dos nucleotidofosfohidrolasas (44) y una topoisomerasa sensible a cumermicina A1 (45). Actividades enzimáticas similares se encuentran también en el virus de la vacuna (46,47).

#### 2.4. DNA.

El genoma del virus de la PPA es un DNA lineal de doble cadena cuya longitud de contorno es aproximadamente 58 nm y cuyo coeficiente de sedimentación tiene un valor 60S (48-50). La molécula reasocia con un valor  $(Cot)^{1/2}$  de 0.60 mol<sup>-1</sup>s<sup>-1</sup>. Estos valores indican que la masa molar del DNA del VPPA es aproximadamente  $100 \times 10^6$  mol<sup>-1</sup> (170 kpb). Su densidad en CsCl es 1.700 g cm<sup>-3</sup>, lo que corresponde a un contenido de G+C del 41% (50).

El estudio de las propiedades hidrodinámicas del DNA del VPPA por ultracentrifugación en gradientes alcalinos de sacarosa y de la renaturalización del DNA no fragmentado sugerían la existencia de un pequeño número de uniones covalentes entre las dos cadenas del DNA localizadas en los extremos de la molécula (51).

Además, si se fragmentaba el DNA del VPPA por incubación en presencia de una endonucleasa de restricción y los productos se desnaturalizan y renaturalizan rápidamente, solamente los fragmentos terminales se recuperan como moléculas bicatenarias resistentes a la acción de la nucleasa S1, siendo identificables por electroforesis en gel (52). Todos

Otras actividades presentes son una proteínquinasa (35), una o posiblemente dos nucleotidofosfohidrolasas (44) y una topoisomerasa sensible a cumermicina A1 (45). Actividades enzimáticas similares se encuentran también en el virus de la vacuna (46,47).

#### 2.4. DNA.

El genoma del virus de la PPA es un DNA lineal de doble cadena cuya longitud de contorno es aproximadamente 58 nm y cuyo coeficiente de sedimentación tiene un valor 60S (48-50). La molécula reasocia con un valor  $(Cot)_{1/2}$  de 0.60 mol<sup>-1</sup>s<sup>-1</sup>. Estos valores indican que la masa molar del DNA del VPPA es aproximadamente  $100 \times 10^6$  mol<sup>-1</sup> (170 kpb). Su densidad en CsCl es 1.700 g cm<sup>-3</sup>, lo que corresponde a un contenido de G+C del 41% (50).

El estudio de las propiedades hidrodinámicas del DNA del VPPA por ultracentrifugación en gradientes alcalinos de sacarosa y de la renaturalización del DNA no fragmentado sugerían la existencia de un pequeño número de uniones covalentes entre las dos cadenas del DNA localizadas en los extremos de la molécula (51).

Además, si se fragmentaba el DNA del VPPA por incubación en presencia de una endonucleasa de restricción y los productos se desnaturalizan y renaturalizan rápidamente, solamente los fragmentos terminales se recuperan como moléculas bicatenarias resistentes a la acción de la nucleasa S1, siendo identificables por electroforesis en gel (52). Todos

estos datos indican que el DNA del VPPA contiene uniones terminales covalentes probablemente semejantes a las presentes en el genoma de los poxvirus (53,54) que se describirán en detalle más adelante.

El genoma del VPPA adaptado a crecer en células VERO (55) ha sido clonado en los plásmidos de *E. coli* pBR325 y pBR322 y en el vector derivado del fago  $\lambda$ ,  $\lambda$ WES $\lambda$ B. Asimismo se ha elaborado un mapa de restricción de dicho genoma en el que se detallan el orden y distancias entre 81 sitios de restricción reconocidos por cinco nucleasas diferentes (SalI, EcoRI, KpnI, PvuI y SmaI) (52). Dicho mapa se muestra en la figura 1.

La hibridación recíproca de los fragmentos terminales indicó la existencia de repeticiones en los extremos de la molécula del DNA (56). Un análisis por microscopía electrónica de los heteroduplex formados por renaturalización de los fragmentos de restricción terminales desnaturalizados y a los que se les habían eliminado los enlaces covalentes terminales demostró la existencia en el genoma de repeticiones terminales invertidas de una longitud de  $2.09 \pm 0.21$  kpb. En condiciones de hibridación menos restrictivas se detectó la presencia de una repetición invertida interna de  $130 \pm 60$  pb de carácter imperfecto, y separada de la repetición terminal por distancias distintas en ambos extremos ( $2.45 \pm 0.38$  kpb en el extremo izquierdo y  $0.44 \pm 0.08$  kpb en el extremo derecho)(143).

El procedimiento de obtención de los clones correspondientes a los fragmentos terminales EcoRI hizo uso de la nucleasa S1 con el fin de eliminar la unión covalente terminal que imposibilita el clonaje en los vectores empleados. Este tratamiento fue la causa de las deleciones observadas en los fragmentos terminales clonados: aproximadamente 1.0 kpb en el fragmento EcoRI-K' y 1.7 kpb en el EcoRI-D'. La adscripción de la proporción de esta deleción a cada extremo del fragmento EcoRI-D' clonado se llevó a cabo por mapeo de restricción de sitios SmaI, concluyéndose que el recombinante poseía deleciones de 1.4 y 0.3 kpb en el extremo izquierdo y derecho, respectivamente (55).

Por convención el genoma se orienta de forma que el extremo izquierdo corresponde al fragmento terminal EcoRI-K' y el derecho al EcoRI-D'.

### 3. INTERACCION VIRUS-CELULA.

#### 3.1. Penetración.

No existen datos experimentales sobre la adsorción y penetración del VPPA en su célula blanco natural, los macrófagos porcinos. En el caso del VPPA adaptado a células VERO, la infección es inhibida de forma reversible por drogas lisosomotrópicas, lo que sugiere que la entrada del virus ocurre por un mecanismo de endocitosis mediada por receptor (56).



El procedimiento de obtención de los clones correspondientes a los fragmentos terminales EcoRI hizo uso de la nucleasa S1 con el fin de eliminar la unión covalente terminal que imposibilita el clonaje en los vectores empleados. Este tratamiento fue la causa de las deleciones observadas en los fragmentos terminales clonados: aproximadamente 1.0 kpb en el fragmento EcoRI-K' y 1.7 kpb en el EcoRI-D'. La adscripción de la proporción de esta deleción a cada extremo del fragmento EcoRI-D' clonado se llevó a cabo por mapeo de restricción de sitios SmaI, concluyéndose que el recombinante poseía deleciones de 1.4 y 0.3 kpb en el extremo izquierdo y derecho, respectivamente (55).

Por convención el genoma se orienta de forma que el extremo izquierdo corresponde al fragmento terminal EcoRI-K' y el derecho al EcoRI-D'.

### 3. INTERACCION VIRUS-CELULA.

#### 3.1. Penetración.

No existen datos experimentales sobre la adsorción y penetración del VPPA en su célula blanco natural, los macrófagos porcinos. En el caso del VPPA adaptado a células VERO, la infección es inhibida de forma reversible por drogas lisosomotrópicas, lo que sugiere que la entrada del virus ocurre por un mecanismo de endocitosis mediada por receptor (56).

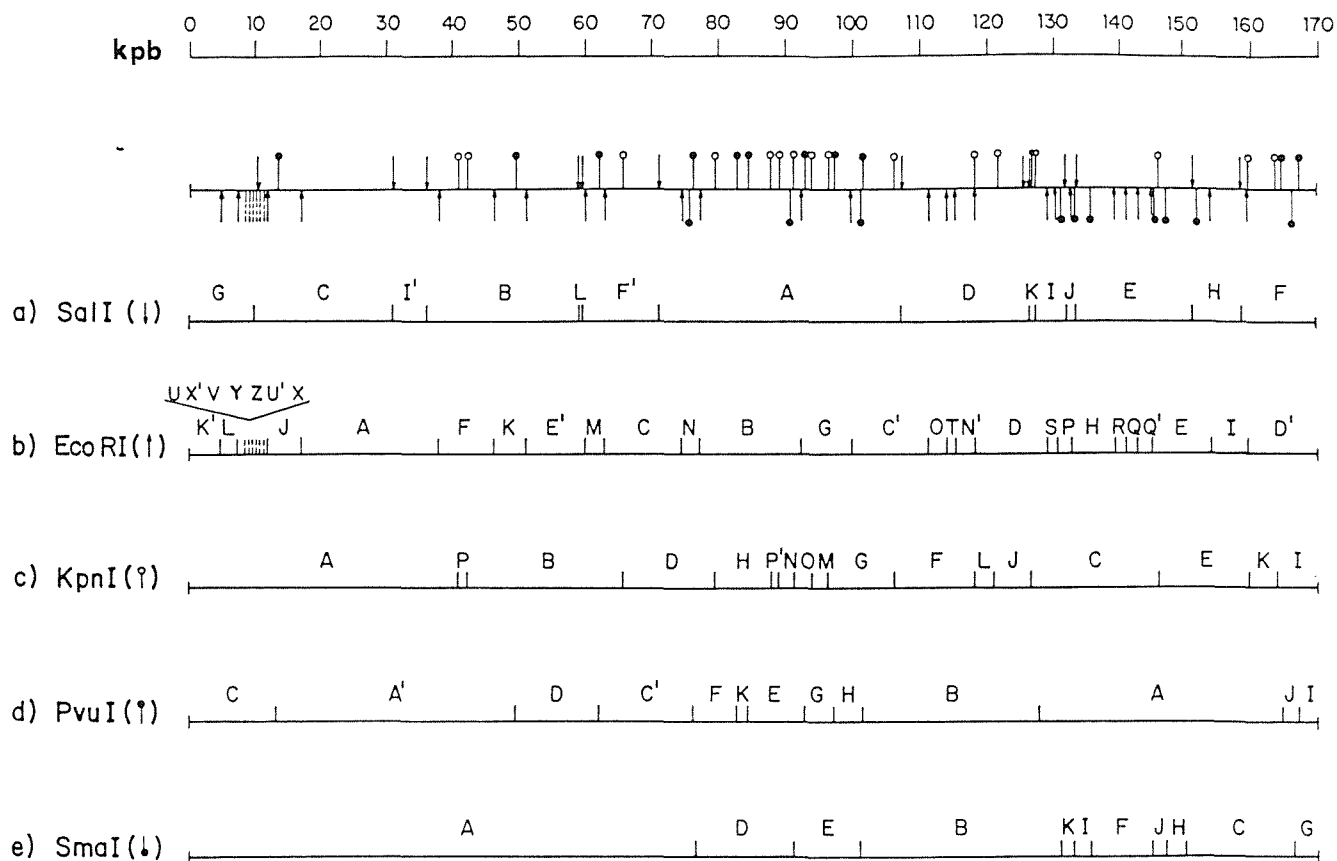


Figura 1. Mapas de restricción del DNA de VPPA, para los enzimas SalI, EcoRI, KpnI, PvuI y SmaI. Los fragmentos se denotan, de mayor a menor tamaño, por letras mayúsculas. Se muestra, en la parte superior, un mapa de restricción compuesto (52).

### 3.2. Síntesis macromolecular.

Tras la penetración del virus, el DNA es transcrito por la RNA-polimerasa del virión, dando lugar a especies de RNA cuya localización en el genoma ha sido determinada por estudios de hibridación (57). Los resultados se muestran en la figura 2. Los RNAs sintetizados en las células infectadas en presencia de cicloheximida o arabinósido de citosina hibridan con las mismas regiones del DNA con las que lo hace el RNA sintetizado por la polimerasa del virión. Estos resultados sugieren que, antes de la replicación del DNA, los RNAs virales son sintetizados por la RNA-polimerasa presente en las partículas virales.

Tras la replicación del DNA aparecen nuevas especies de RNA que hibridan con regiones del DNA que no se han transcrito en las células infectadas en presencia de inhibidores de la síntesis de proteínas o DNA (figura 2). La RNA-polimerasa II no está implicada en la síntesis de ninguno de los dos tipos de RNAs, puesto que las células infectadas en presencia de 5,6-dicloro- $\beta$ -D-ribofuranosilbenzimidazol sintetizan RNAs y proteínas que son indistinguibles de los producidos en ausencia de la droga (58).

Por medio de experimentos de traducción "in vitro" de RNAs seleccionados con fragmentos de restricción clonados se ha confeccionado un mapa de traducción que se muestra asimismo en la figura 2.

La síntesis del DNA viral requiere una DNA-polimerasa inducida por el virus (59) que es sensible al ácido fosfonoacético (60) y un núcleo funcional de la célula infectada (61). Experimentos de pulso y caza realizados en monocitos porcinos infectados indican que la síntesis de DNA viral tiene lugar en el citoplasma (62).

#### 4. ESTRUCTURA, REPLICACION Y EXPRESION DEL GENOMA DE LOS ORTOPOXVIRUS.

Los genomas de los ortopoxvirus consisten en una molécula lineal de DNA bicatenario de un tamaño de aproximadamente 200 kpb cuyos extremos se encuentran unidos covalentemente (53,54). Constan de tres regiones diferenciadas: una región central altamente conservada cuyo tamaño equivale aproximadamente a la mitad del genoma y dos regiones flanqueantes de tamaño similar que han divergido mucho más (63). En los extremos de estas regiones existen repeticiones terminales invertidas, salvo en el caso del virus de la viruela humana (64).

Las RTI del genoma de los ortopoxvirus han sido estudiadas en detalle en el virus vacunal y en el de la viruela bovina ("cowpox") (54,65-67). En estos virus su tamaño es de aproximadamente 10 kpb y comprende dos regiones: una región interna de 7.5 kpb que incluye secuencias codificantes y una región terminal que contiene generalmente dos bloques de secuencias repetidas en tandem muy relacionadas entre sí. El

La síntesis del DNA viral requiere una DNA-polimerasa inducida por el virus (59) que es sensible al ácido fosfonoacético (60) y un núcleo funcional de la célula infectada (61). Experimentos de pulso y caza realizados en monocitos porcinos infectados indican que la síntesis de DNA viral tiene lugar en el citoplasma (62).

#### 4. ESTRUCTURA, REPLICACION Y EXPRESION DEL GENOMA DE LOS ORTOPOXVIRUS.

Los genomas de los ortopoxvirus consisten en una molécula lineal de DNA bicatenario de un tamaño de aproximadamente 200 kpb cuyos extremos se encuentran unidos covalentemente (53,54). Constan de tres regiones diferenciadas: una región central altamente conservada cuyo tamaño equivale aproximadamente a la mitad del genoma y dos regiones flanqueantes de tamaño similar que han divergido mucho más (63). En los extremos de estas regiones existen repeticiones terminales invertidas, salvo en el caso del virus de la viruela humana (64).

Las RTI del genoma de los ortopoxvirus han sido estudiadas en detalle en el virus vacunal y en el de la viruela bovina ("cowpox") (54,65-67). En estos virus su tamaño es de aproximadamente 10 kpb y comprende dos regiones: una región interna de 7.5 kpb que incluye secuencias codificantes y una región terminal que contiene generalmente dos bloques de secuencias repetidas en tandem muy relacionadas entre sí. El

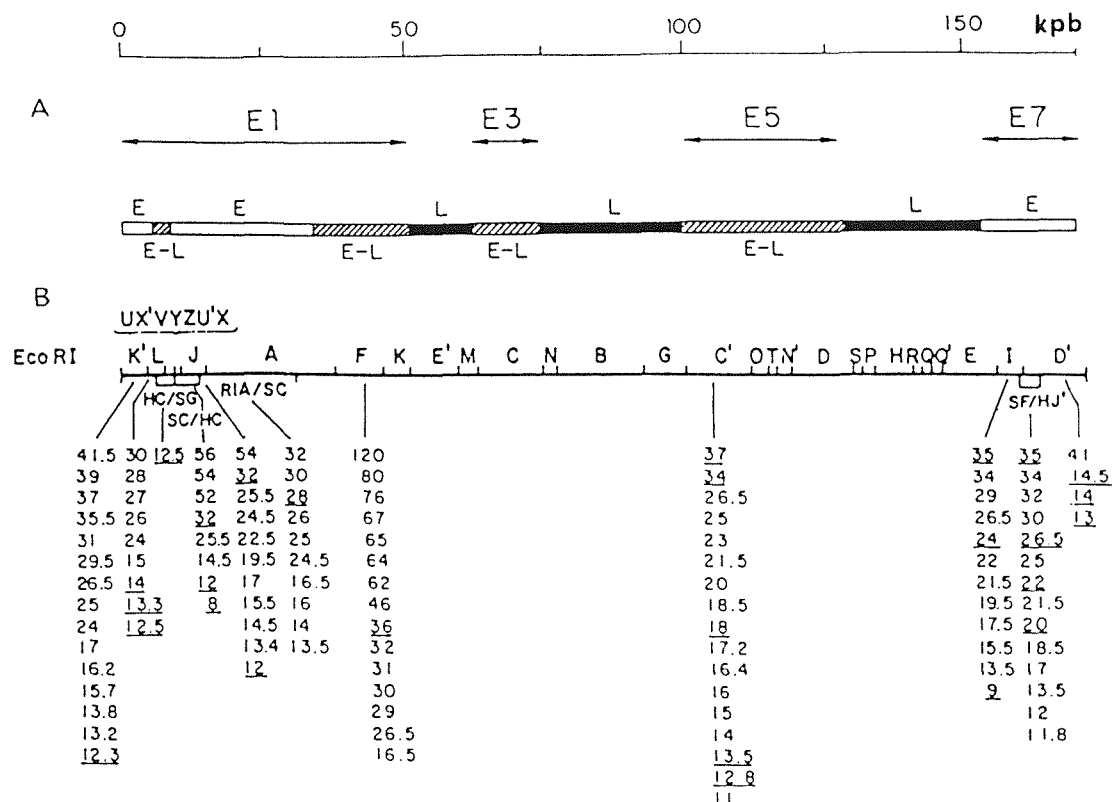


Figura 2. Mapa de transcripción -traducción del DNA de VPPA. A) Mapa de transcripción. E, zonas de transcripción temprana; L, zonas de transcripción tardía; E-L, zonas de transcripción temprana-tardía. B) Mapa de proteínas tempranas del VPPA. Las cifras corresponden a tamaño molecular de las proteínas en K. Los fragmentos de restricción utilizados para la selección han sido EcoRI, HindIII-SalI (HC/SG, SC/HC, SF/HJ') o EcoRI-SalI (RA/SC). El subrayado indica las proteínas presentes en mayor proporción en la selección-traducción con un fragmento específico de DNA (57).

primer bloque se encuentra a una distancia de aproximadamente 150 pb del extremo en ambos casos.

En el caso del virus vacunal este primer bloque consiste en 13 repeticiones en tandem de 70 nucleótidos, mientras que el segundo, separado del primero por una secuencia única de 325 pb de longitud, consta de grupos de repeticiones en tandem de 70, 125 y 54 nucleótidos (repetidos 13, 2 y 8 veces, respectivamente). Las secuencias repetidas en tandem muestran microheterogeneidad, y existe una homología de secuencia considerable entre las repeticiones de distinta longitud (65,68).

La RTI presente en el genoma del poxvirus bovino muestra idéntica organización general, constando los grupos de repeticiones en tandem de cuatro tipos básicos de subunidades de aproximadamente 50 nucleótidos. Las subunidades muestran una homología considerable entre sí y con las secuencias repetidas en tandem del virus vacunal (66).

El primer grupo consta de un total de 36 repeticiones que a su vez se pueden agrupar en subgrupos de un corto número de subunidades (de 2 a 4), con un total de 1502 nucleótidos. Separado de este grupo por una secuencia única de 312 nucleótidos se encuentra el segundo grupo, de menor tamaño (13 subunidades y 547 nucleótidos totales), en los que se pueden definir subgrupos de 3 subunidades.

Tanto la región única que separa el extremo del genoma del primer grupo de repeticiones en tandem, como la que separa ambos grupos, presentan una secuencia casi idéntica en

los genomas de ambos virus, con una homología superior al 95%.

La naturaleza exacta de la unión covalente terminal entre ambas cadenas del genoma de los ortopoxvirus solo ha sido determinada por secuenciación en el caso del virus de la vacuna (54). Consiste en una horquilla terminal, con lo que la molécula lineal del DNA bicatenario constituye una única cadena continua. Existen asimismo datos experimentales consistentes con la presencia de horquillas terminales en otros miembros de la familia de los poxvirus, incluyendo poxvirus de aves (69), molluscum contagiosum (70) y el virus de la stomatitis papullosa (71).

La horquilla terminal existe en dos formas, presentes en cantidades aproximadamente equimolares en cada extremo del genoma y cuya existencia se puso de manifiesto por la existencia de dos tipos de fragmentos terminales de restricción, de igual tamaño, pero distinguibles por su movilidad electroforética en condiciones no desnaturalizantes.

La determinación de la secuencia de ambos tipos de fragmentos terminales utilizando el método de Maxam y Gilbert (72) reveló que el vértice de la horquilla terminal está constituido por 104 nucleótidos con una composición muy rica en A+T (92%) y que no pueden aparear completamente (16 bases desapareadas en la configuración de máximo apareamiento). Las secuencias de ambas formas son diferentes, pero si se excluyen las bases desapareadas son idénticas. Además, si se comparan ambas secuencias escritas de forma lineal y en



los genomas de ambos virus, con una homología superior al 95%.

La naturaleza exacta de la unión covalente terminal entre ambas cadenas del genoma de los ortopoxvirus solo ha sido determinada por secuenciación en el caso del virus de la vacuna (54). Consiste en una horquilla terminal, con lo que la molécula lineal del DNA bicatenario constituye una única cadena continua. Existen asimismo datos experimentales consistentes con la presencia de horquillas terminales en otros miembros de la familia de los poxvirus, incluyendo poxvirus de aves (69), molluscum contagiosum (70) y el virus de la stomatitis papulosa (71).

La horquilla terminal existe en dos formas, presentes en cantidades aproximadamente equimolares en cada extremo del genoma y cuya existencia se puso de manifiesto por la existencia de dos tipos de fragmentos terminales de restricción, de igual tamaño, pero distinguibles por su movilidad electroforética en condiciones no desnaturalizantes.

La determinación de la secuencia de ambos tipos de fragmentos terminales utilizando el método de Maxam y Gilbert (72) reveló que el vértice de la horquilla terminal está constituido por 104 nucleótidos con una composición muy rica en A+T (92%) y que no pueden aparear completamente (16 bases desapareadas en la configuración de máximo apareamiento). Las secuencias de ambas formas son diferentes, pero si se excluyen las bases desapareadas son idénticas. Además, si se comparan ambas secuencias escritas de forma lineal y en

orientaciones opuestas resultan ser invertidas y complementarias entre sí (69).

Los modelos de replicación del genoma de los ortopoxvirus tienen que dar cuenta de dos observaciones experimentales: una de ellas es la existencia de las horquillas terminales presentes en sus dos formas en ambos extremos del genoma, la otra es la presencia de intermediarios replicativos en forma de concatémeros cabeza-cabeza o cola-cola (68,73).

Los esquemas elaborados son similares a los propuestos por Cavalier-Smith (74) y Bateman (75) para la replicación de los cromosomas eucarióticos y del genoma de los parvovirus (76). Se basan en la transición entre estructuras palindrómicas y de horquilla de los extremos del genoma, gracias a la acción de un conjunto de enzimas cuya existencia se postula.

Los modelos se dividen en dos tipos: el primero supone una autoiniciación de la replicación, tras la generación de una interrupción en la cadena polinucleotídica, a la que sigue una transición horquilla-palíndrome por replicación a partir del hidroxilo generado en 3' (68,71). El segundo supone un inicio "de novo" mediado por iniciadores, posiblemente moléculas cortas de RNA (77).

En la figura 3 se representan dos posibilidades dentro del modelo de autoiniciación, según que se produzcan uno o dos interrupciones de la cadena por genoma. En el primer caso, tras la generación del palíndrome, se vuelven a formar dos horquillas en un extremo continuándose la replicación

elongando el extremo 3' y desplazando a la cadena parental, con lo que se pueden generar concatémeros cabeza-cabeza y cola-cola de tamaño dímero o mayor. En la segunda posibilidad, se generan palíndromes y se regeneran horquillas en ambos extremos del genoma, prosiguiendo la replicación bidireccionalmente, con lo que se obtienen dos moléculas hijas.

La generación de moléculas hijas con cadenas ininterrumpidas supondría, partiendo de los concatémeros, un proceso de roturas de la cadena que dieran lugar a moléculas de tamaño monómero y otro proceso de ligamiento de la cadena polinucleotídica previa reformatión de la horquilla terminal, como se muestra en dicha figura 3. Solo se necesitaría este último paso de ligamiento en la alternativa de replicación bidireccional.

En el caso del modelo de iniciación "de novo", el primer paso no es la generación de una interrupción de cadena, sino una elongación de un iniciador, unido al genoma no necesariamente en la zona de la horquilla terminal, como se muestra en la figura 3. La generación de moléculas hijas implica pasos de generación de interrupciones en la cadena, regeneración de la horquilla terminal y ligamiento para dar lugar a cadenas ininterrumpidas.

Ambos tipos de modelos posibilitan tanto la formación de concatémeros como la obtención de moléculas hijas que poseen todas las combinaciones posibles de horquillas terminales en sus extremos.

elongando el extremo 3' y desplazando a la cadena parental, con lo que se pueden generar concatémeros cabeza-cabeza y cola-cola de tamaño dímero o mayor. En la segunda posibilidad, se generan palíndromes y se regeneran horquillas en ambos extremos del genoma, prosiguiendo la replicación bidireccionalmente, con lo que se obtienen dos moléculas hijas.

La generación de moléculas hijas con cadenas ininterrumpidas supondría, partiendo de los concatémeros, un proceso de roturas de la cadena que dieran lugar a moléculas de tamaño monómero y otro proceso de ligamiento de la cadena polinucleotídica previa reformación de la horquilla terminal, como se muestra en dicha figura 3. Solo se necesitaría este último paso de ligamiento en la alternativa de replicación bidireccional.

En el caso del modelo de iniciación "de novo", el primer paso no es la generación de una interrupción de cadena, sino una elongación de un iniciador, unido al genoma no necesariamente en la zona de la horquilla terminal, como se muestra en la figura 3. La generación de moléculas hijas implica pasos de generación de interrupciones en la cadena, regeneración de la horquilla terminal y ligamiento para dar lugar a cadenas ininterrumpidas.

Ambos tipos de modelos posibilitan tanto la formación de concatémeros como la obtención de moléculas hijas que poseen todas las combinaciones posibles de horquillas terminales en sus extremos.

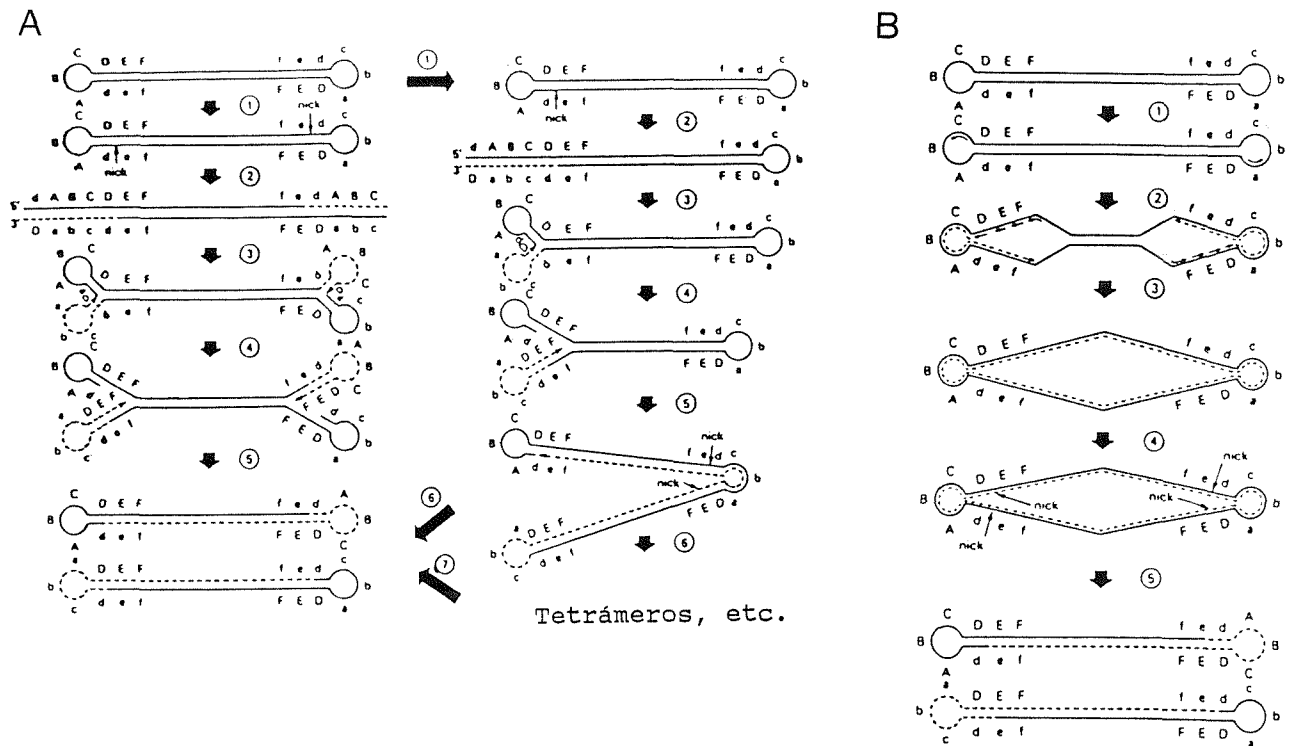


Figura 3. Modelos de replicación del DNA del virus de la vacuna. Las secuencias complementarias se denotan con letras mayúsculas y minúsculas. Las horquillas teerminales están dibujadas como lazos (ABC,abc). La repetición terminal invertida se denota como DEF, def. A) Modelos de autoiniciación. La iniciación puede ocurrir por rotura ("nicks") de una de las cadenas del DNA en un extremo (derecha) o en ambos extremos simultáneamente (izquierda). Solo en el segundo caso se originarían concatémeros. B) Modelo de iniciación "de novo". El comienzo de la síntesis vía RNAs iniciadores ocurre simultáneamente en ambas horquillas terminales. Si la iniciación tuviera lugar en un único extremo o próximo al mismo, los modelos resultantes serían similares al representado. Tras el paso (3) se obtendría un dímero circular(68).

Los datos experimentales disponibles no permiten discriminar entre los dos tipos de modelos propuestos. Existen datos que sugieren que se producen interrupciones en la cadena de la molécula parental en posiciones próximas a los extremos del genoma (78,79), lo que apoyaría el modelo de autoiniciación. Sin embargo estudios adicionales indicaron que la replicación comienza cerca de los extremos de la molécula, siendo de tipo semiconservativo e implicando síntesis discontinua del DNA y el uso de moléculas de RNA iniciadoras (77,80-82), lo que sería consistente con el modelo de iniciación "de novo". El ligamiento de las interrupciones de la cadena polinucleotídica ocurre relativamente tarde en el curso de la infección (77,78), hecho que está en consonancia con ambos tipos de modelos.

Estos mecanismos puede que sean aplicables no solo a la replicación del genoma de los ortopoxvirus, sino a la de los cromosomas eucarióticos cuyos telómeros pueden tener estructura de horquilla, como sugiere la existencia de uniones covalentes terminales hallada en los cromosomas de levadura (83).

En cuanto a la expresión del genoma de los ortopoxvirus, y centrándose en su representante más estudiado, el virus de la vacuna, se puede afirmar que el programa transcripcional de este virus es especial puesto que los enzimas implicados en la síntesis y el procesamiento del RNA, mensajero son similares a las de las células eucarióticas y sin

embargo se encuentran empaquetados en la partícula viral madura y están codificadas por el virus.

Se han mapeado de forma precisa muchos genes tempranos y tardíos y algunos de ellos se han secuenciado, incluyendo en estos últimos un gen tardío y los tres genes tempranos (7.5, 19 y 42 K) codificados en la RTI (84-89). No se ha encontrado en ningún caso evidencia de procesos de maduración del RNA que impliquen cortes y eliminación de secuencias, de moléculas precursoras del RNA mensajero maduro ("splicing"). Los RNAs mensajeros tardíos parecen ser de mayor longitud que los tempranos e incluir secuencias tempranas (89). Presentan asimismo la característica de ser extremadamente heterogéneos en tamaño, y de exceder en algunas ocasiones su longitud varias veces a la necesaria para la codificación de la proteína (90-92). La heterogeneidad de tamaño se debe a sus extremos 3' (88). Tanto en genes tardíos como en tempranos se han encontrado ejemplos de transcritos solapantes (93,94).

En cuanto a la presencia de señales reguladoras de la transcripción en la secuencia adyacente en dirección 5' a la correspondiente al inicio del RNA mensajero, no se han encontrado secuencias homólogas a las consenso de reconocimiento de la RNA polimerasa II eucariótica. El análisis comparativo de dichas secuencias revela sin embargo rasgos comunes: un alto contenido en A+T y la presencia en posiciones -13 y -40 de las secuencias AA(A/T)AATA(A/T)A y TAT-AAT(T/A)A, respectivamente (87,88,95), en el caso de los ge-

embargo se encuentran empaquetados en la partícula viral madura y están codificadas por el virus.

Se han mapeado de forma precisa muchos genes tempranos y tardíos y algunos de ellos se han secuenciado, incluyendo en estos últimos un gen tardío y los tres genes tempranos (7.5, 19 y 42 K) codificados en la RTI (84-89). No se ha encontrado en ningún caso evidencia de procesos de maduración del RNA que impliquen cortes y eliminación de secuencias, de moléculas precursoras del RNA mensajero maduro ("splicing"). Los RNAs mensajeros tardíos parecen ser de mayor longitud que los tempranos e incluir secuencias tempranas (89). Presentan asimismo la característica de ser extremadamente heterogéneos en tamaño, y de exceder en algunas ocasiones su longitud varias veces a la necesaria para la codificación de la proteína (90-92). La heterogeneidad de tamaño se debe a sus extremos 3' (88). Tanto en genes tardíos como en tempranos se han encontrado ejemplos de transcritos solapantes (93,94).

En cuanto a la presencia de señales reguladoras de la transcripción en la secuencia adyacente en dirección 5' a la correspondiente al inicio del RNA mensajero, no se han encontrado secuencias homólogas a las consenso de reconocimiento de la RNA polimerasa II eucariótica. El análisis comparativo de dichas secuencias revela sin embargo rasgos comunes: un alto contenido en A+T y la presencia en posiciones -13 y -40 de las secuencias AA(A/T)AATA(A/T)A y TAT-AAT(T/A)A, respectivamente (87,88,95), en el caso de los ge-



nes tempranos estudiados (7.5, 19 y 42 K y timidina kinasas). El gen tardío 28 K presenta en posición -13 la secuencia Ag (96). Asimismo se ha detectado la presencia de palíndromos en estas regiones (97).

En cuanto a las señales de terminación de la transcripción, no se han detectado secuencias consenso. Sin embargo, en algunos casos existen repeticiones cortas en tandem próximas a los extremos 3' del RNA mensajero: (CTATT/CC)<sub>2</sub> en el gen temprano 7.5 K y una secuencia de 11 nucleótidos TAGAGG-TAGAGG en el caso del gen 42 K. Inmediatamente antes del final del RNA mensajero del gen 19 K (como los dos anteriores, codificado en la RTI) se encuentra la secuencia CTTATG, relacionada con la presente en el gen 7.5 K. En ningún caso se ha encontrado la señal canónica de poliadenilación (AATAAA).

En cuanto a la presencia de secuencias implicadas en la unión al ribosoma de los RNAs mensajeros por complementariedad al RNA ribosomal 18S, solo se han encontrado en los genes 19 K (ATCCATC) y 7.5 K (TCTT), pero no en el 42 K (87).

Las secuencias iniciales de los RNAs mensajeros estudiados muestran un significativo grado de homología con la secuencia consenso eucariótica de iniciación CC(A/G)CCAUG(G) definida estadísticamente (98).

Como conclusión de estos análisis, parece evidente que el virus de la vacuna utiliza señales reguladoras únicas, como parecen apoyar asimismo estudios transcripcionales rea-

lizados con extractos preparados a partir de células infectadas por el virus y no infectadas (99).

## 5. TECNOLOGIA DE LA SECUENCIACION DE DNA.

Las técnicas de secuenciación rápida de DNA se basan en la generación de una mezcla de oligonucleótidos que se extienden desde un extremo fijo hasta alguna de las cuatro bases posibles en la cadena del DNA. Tras generar una colección similar de oligonucleótidos marcados radiactivamente correspondiente a cada una de las cuatro bases, se resuelven dichas mezclas por tamaño mediante electroforesis de alta resolución en geles de poliacrilamida. La secuencia de DNA se puede leer directamente de la autorradiografía de tales oligonucleótidos así resueltos.

La obtención de las mezclas de oligonucleótidos se denomina reacciones de secuenciación. La preparación de los distintos fragmentos de DNA cuya secuencia será determinada total o parcialmente gracias a las reacciones y a la electroforesis posterior, obteniéndose de este modo la secuencia global de la zona de interés, se denomina estrategia de secuenciación.

Los dos tipos de reacciones más utilizados son la degradación química específica de Maxam y Gilbert (72) y la síntesis enzimática interrumpida por dideoxinucleótidos de Sanger y col. (100).

lizados con extractos preparados a partir de células infectadas por el virus y no infectadas (99).

## 5. TECNOLOGIA DE LA SECUENCIACION DE DNA.

Las técnicas de secuenciación rápida de DNA se basan en la generación de una mezcla de oligonucleótidos que se extienden desde un extremo fijo hasta alguna de las cuatro bases posibles en la cadena del DNA. Tras generar una colección similar de oligonucleótidos marcados radiactivamente correspondiente a cada una de las cuatro bases, se resuelven dichas mezclas por tamaño mediante electroforesis de alta resolución en geles de poliacrilamida. La secuencia de DNA se puede leer directamente de la autorradiografía de tales oligonucleótidos así resueltos.

La obtención de las mezclas de oligonucleótidos se denomina reacciones de secuenciación. La preparación de los distintos fragmentos de DNA cuya secuencia será determinada total o parcialmente gracias a las reacciones y a la electroforesis posterior, obteniéndose de este modo la secuencia global de la zona de interés, se denomina estrategia de secuenciación.

Los dos tipos de reacciones más utilizados son la degradación química específica de Maxam y Gilbert (72) y la síntesis enzimática interrumpida por dideoxinucleótidos de Sanger y col. (100).

Las reacciones químicas específicas de una o dos de las cuatro bases implican generalmente tres pasos consecutivos: modificación de una base; eliminación de la base modificada del esqueleto hidrocarbonado y escisión de dicho esqueleto en la posición correspondiente a la base eliminada. De entre las muchas reacciones desarrolladas, las más utilizadas son las específicas de G, G+A, T+C, C, descritas por Maxam y Gilbert (72).

En cuanto a las reacciones enzimáticas, en la actualidad el método mas utilizado implica el uso del fragmento "Klenow" de la DNA polimerasa I de E. coli, de moldes generados por clonaje en los vectores derivados del bacteriófago de cadena sencilla M13 de los fragmentos a secuenciar y de un iniciador universal de la síntesis consistente en un fragmento de DNA homólogo a las secuencias de M13 adyacentes al sitio de inserción.

En la técnica de Maxam y Gilbert, el extremo fijo en la mezcla de oligonucleótidos se genera por un enzima de restricción y los extremos variables específicos de cada base por la reacciones químicas de degradación. Los oligonucleótidos separados por electroforesis se visualizan gracias a la marca terminal radiactiva introducida en el extremo constante.

En cambio, en la técnica de síntesis en presencia de terminadores el extremo fijado es el de un iniciador, los extremos variables específicos se generan al introducirse en

la cadena naciente dideoxinucleótidos, y la marca radiactiva se incorpora durante la misma síntesis.

La electroforesis emplea geles de acrilamida-bisacrilamida muy finos (0.1-0.5 mm), que contienen urea como agente desnaturalizante, lo que unido a la elevada temperatura a la que se lleva a cabo, (55-60 °C) consigue que la separación de los oligonucleótidos se lleve a cabo exclusivamente por su tamaño. Se logran así resolver diferencias de un nucleótido en fragmentos de hasta 500 nucleótidos, dependiendo de las dimensiones y concentración del gel.

Los últimos avances que aumentan el poder de resolución de estas electroforesis son el uso del isótopo  $^{35}\text{S}$  en lugar del  $^{32}\text{P}$  (101) y la transferencia a una membrana rotatoria de los productos separados por el gel (102).

Las primeras estrategias de secuenciación utilizadas para generar lecturas solapantes se basaban en la existencia de un mapa de restricción detallado de la zona de interés, química o utilizables como iniciadores en las reacciones de síntesis enzimática. Las estrategias más recientes requieren muy poca, o ninguna, información sobre los sitios de restricción presentes en la zona a secuenciar.

Una de ellas, utilizada generalmente combinada con las reacciones enzimáticas, consiste en la obtención de colecciones de subfragmentos generados al azar y clonaje en vectores derivados de M13. El solapamiento de las lecturas de un número abundante de clones dará como resultado la secuencia global. Se han utilizado para generar los subfragmentos:

la cadena naciente dideoxinucleótidos, y la marca radiactiva se incorpora durante la misma síntesis.

La electroforesis emplea geles de acrilamida-bisacrilamida muy finos (0.1-0.5 mm), que contienen urea como agente desnaturalizante, lo que unido a la elevada temperatura a la que se lleva a cabo, (55-60 °C) consigue que la separación de los oligonucleótidos se lleve a cabo exclusivamente por su tamaño. Se logran así resolver diferencias de un nucleótido en fragmentos de hasta 500 nucleótidos, dependiendo de las dimensiones y concentración del gel.

Los últimos avances que aumentan el poder de resolución de estas electroforesis son el uso del isótopo  $^{35}\text{S}$  en lugar del  $^{32}\text{P}$  (101) y la transferencia a una membrana rotatoria de los productos separados por el gel (102).

Las primeras estrategias de secuenciación utilizadas para generar lecturas solapantes se basaban en la existencia de un mapa de restricción detallado de la zona de interés, química o utilizables como iniciadores en las reacciones de síntesis enzimática. Las estrategias más recientes requieren muy poca, o ninguna, información sobre los sitios de restricción presentes en la zona a secuenciar.

Una de ellas, utilizada generalmente combinada con las reacciones enzimáticas, consiste en la obtención de colecciones de subfragmentos generados al azar y clonaje en vectores derivados de M13. El solapamiento de las lecturas de un número abundante de clones dará como resultado la secuencia global. Se han utilizado para generar los subfragmentos:

enzimas de restricción con secuencia de reconocimiento de cuatro bases, sonicación o tratamiento con DNasa I. Los dos primeros métodos se han aplicado a grandes proyectos de secuenciación, como son los de los genomas del bacteriófago  $\lambda$  y del virus de Epstein-Barr (103,104).

Otras estrategias que combinan la generación de clones al azar con otras aproximaciones dirigidas son, según la revisión de Deininger (105): a) la secuenciación utilizando una sola reacción específica para identificar y descartar clones idénticos (106); b) la resecuenciación a partir de clones ya utilizados empleando condiciones de electroforesis que permitan la lectura de un mayor número de bases; c) la secuenciación desde el otro extremo de determinados clones, bien por reclonaje del inserto en la orientación contraria (107) o por utilización de otro iniciador previa síntesis de inserto bicatenario; d) clonaje de fragmentos generados por enzimas de restricción; e) selección de clones presentes en otra colección generada al azar por hibridación con los clones de zonas adyacentes a aquellos cuya secuencia queda por determinar (103).

Las estrategias dirigidas se agrupan en tres grandes tipos. El primero de ellos es aplicable solo a la secuenciación por degradación química y utiliza una combinación de mercuriación en un único extremo, digestión parcial con enzimas de restricción y marcaje de los extremos no mercuriados (108). La selección por cromatografía en tiolagarosa y electroforesis en agarosa permite obtener fragmentos con un

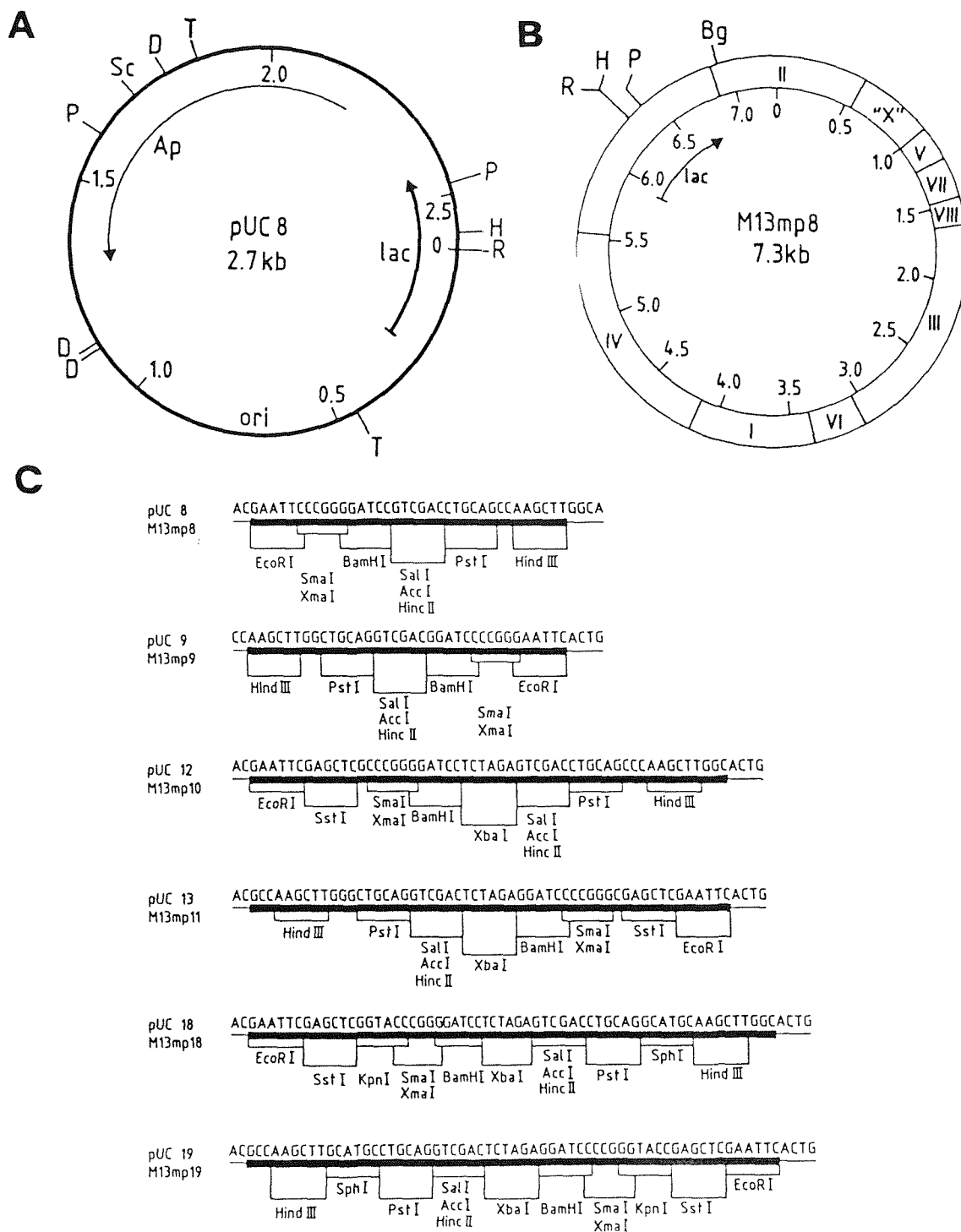
Una serie análoga de plásmidos (pUC) que contienen los sitios de clonaje múltiple en ambas orientaciones respecto al gen lac ha sido desarrollada por los mismos autores (115). Contienen un gen de resistencia a ampicilina modificado procedente de pBR322 y permiten asimismo seleccionar directamente los plásmidos recombinantes en presencia de X-gal.

La serie de vectores M13mp/pUC se ha ampliado incluyendo cada vez más sitios únicos de restricción en el sitio de clonaje múltiple. En una reciente publicación (116) se describen las secuencias completas de los vectores M13mpl8 y pUC19, así como diversas cepas bacterianas utilizadas como receptores en los procedimientos de transfección/transformación.

El mapa de restricción de un vector representativo de cada serie (M13mp8 y pUC8), acompañado de las secuencias de los distintos sitios de clonaje múltiple se muestran en la figura 4.

Los vectores de la serie pUC ofrecen diversas ventajas para su utilización como vectores de subclonaje en las estrategias dirigidas compatibles con las reacciones químicas de secuenciación. Su sitio de clonaje múltiple permite el subclonaje orientado de fragmentos de restricción, o de fragmentos procedentes de digestión progresiva con nucleasas, empleada en estrategias descritas anteriormente. La selección de recombinantes adecuados se facilita con el sistema basado en la actividad  $\beta$ -galactosidasa. El sitio de clo-





**Figura 4. Vectores de clonaje de las series pUC/M13mp.**

A) Mapa de restricción de pUC18. Se indican, además de sitios de restricción no presentes en el sitio de clonaje múltiple, el origen de replicación (ori), el gen de resistencia a ampicilina (Ap) y el segmento del operón lac. De las secuencias de reconocimiento presentes en el SCM, solo se señalan las correspondientes a HindIII y EcoRI.

B) Mapa de restricción de M13mp8. Se indican sitios de restricción y el segmento del operón lac como en A). Los números romanos indican las secuencias codificantes virales.

C) Secuencia de nucleótidos de los SCM de los vectores M13mp/pUC. Se indican asimismo los sitios de restricción (116).

naje múltiple permite asimismo generar fácilmente fragmentos marcados terminalmente en un único extremo utilizando enzimas de restricción y los diversos procedimientos de marcaje terminal. Además, se trata de plásmidos de alto número de copias, lo que facilita su purificación eficiente.

## 6. OBJETIVOS DEL TRABAJO.

Las características estructurales singulares de los extremos del genoma del VPPA, como son la presencia de enlaces covalentes terminales y de repeticiones invertidas, tanto terminales como internas, confieren a esta zona del genoma un elevado interés, especialmente por su posible implicación tanto en la replicación del genoma del virus como por su similitud con estructuras de una relevancia biológica más general, como son los telómeros de los cromosomas eucarióticos.

La elucidación de la estructura de la repetición terminal invertida permitiría analizar las homologías y diferencias existentes con las ya estudiadas de poxvirus, con implicaciones acerca de las relaciones filogenéticas entre el VPPA y el grupo de los poxvirus.

En cuanto a las secuencias adyacentes a la repetición terminal invertida en cada extremo del genoma, la detección de posibles genes ayudaría a la comprensión de la expresión del genoma del VPPA.

Por esta razones, los objetivos del presente trabajo han sido la determinación de la secuencia de nucleótidos de las repeticiones terminales invertidas e internas, de las secuencias adyacentes a la repetición terminal invertida y de la unión covalente terminal entre ambas cadenas del DNA presentes en ambos extremos del genoma.

## MATERIALES

## 1. MATERIAL BIOLOGICO.

### 1.1. DNA.

#### 1.1.1. VPPA.

El DNA del virus de la peste porcina africana adaptado a células VERO (BA71-V) fué extraído de partículas virales purificadas por J.M. Almendral (52).

Los plásmidos recombinantes p5RK'1, p5RD'2, p5RL1 y p8H-H y el DNA del fago recombinante LRD'16, que contienen insertos de DNA del VPPA adaptado a células VERO (55), fueron obtenidos por A. Beloso, V. Ley y A. Talavera, según métodos descritos (117).

#### 1.1.2. Vectores.

Las formas replicativas de los fagos M13mp8 y 9 (113) fueron adquiridas a P.L. Biochemicals.

Los plásmidos pUC8, 9, 12, 13, 18 y 19 (116) fueron suministrados por A. González.

#### 1.1.3. Marcadores de peso molecular.

El DNA del fago digerido con HindIII y el de  $\phi$ X174 (forma replicativa) digerido con HaeIII fueron adquiridos a New England Biolabs.

## 1.2.Bacterias.

En este trabajo se han utilizado las siguientes cepas de E. coli:

-JM83: ara,  $\Delta$  lac-pro, strA, thi,  $\phi$ 80dlacZ AM15 (115).

-JM101: thi, sup E,  $\Delta$  lac-pro, F'traD36, proAB, lac IQZAM15 (113).

-GM119: dcm 6, dam 3, met B1, gal K2, gal T22, lac Y1, thi 1, sup E44, tsX78 (mH1, ton A31)(118).

## 2. MEDIOS DE CULTIVO.

-Medio L: Triptona 10 g/l, extracto de levadura 5 g/l y NaCl 5 g/l.

-El medio sólido se preparó por adición al medio líquido de 15 g/l de agar. El agar de plaqueo se preparó por adición de 6 g/l de agar.

-El medio sólido conteniendo ampicilina y X-gal (L-amp-X-gal) se preparó añadiendo al medio sólido ampicilina hasta una concentración final de 150 mg/l. Una vez solidificadas las placas se extendieron sobre el agar 50  $\mu$ l de solución de X-gal en dimetilformamida (2% p/v).

## 3. PRODUCTOS DE ORIGEN COMERCIAL (Lista por orden alfabético).

Acetato amónico (Merck).

Acetato magnésico (Merck).

Acetato potásico (Merck).

Acetato sódico.(Merck).  
Acido acético (Merck).  
Acido bórico (Merck).  
Acido clorhídrico (Merck).  
Acido fórmico (Merck).  
Acido nitriloacético (Fluka).  
Acrilamida (Serva o Merck).  
Adenosina-5'-trifosfato (Sigma).  
Adenosina-5'-trifosfato- $\gamma$ - $^{32}\text{P}$  (Amersham).  
Agarosa tipo II (Sigma).  
Agarosa de bajo punto de fusión (BioRad).  
Alcohol isoamílico (Merck).  
Ampicilina (Sigma).  
Azul de bromofenol (Schering).  
Bisacrilamida (Serva).  
BSA (Sigma).  
1-Butanol (Merck).  
Cacodilato sódico (Merck).  
Citrato sódico (Merck).  
Cloroformo (Merck).  
Cloruro cálcico (Merck).  
Cloruro de cesio (Sigma).  
Cloruro magnésico (Merck).  
Cloruro sódico (Merck).  
CTAB (Sigma).  
DMS (Aldrich).  
DMSO (Carlo Erba).

Deoxinucleósidos-5'-trifosfato(Sigma o Böhringer Mannheim).

Deoxinucleósidos-5'-trifosfato- $\alpha$ -<sup>32</sup>P (Amersham).

DNA ligasa del fago T4 (New England Biolabs o Böhringer Mannheim).

DNA polimerasa I de E. coli (New England Biolabs o Böhringer Mannheim).

DNA Polimerasa I de E. coli-fragmento "Klenow" (P.L. Biochemicals o Böhringer Mannheim).

DNAts (Sigma).

DNAsa I (Worthington Biochemical).

DTT (Sigma).

EDTA (Merck).

Endonucleasas de restricción (New England Biolabs o Böhringer Mannheim).

Etanol (Merck).

EtBr (Sigma).

Ficoll (Sigma).

Fijador FX40 (Kodak).

Fosfatasa alcalina de intestino de ternera (Böhringer Mannheim).

Fosfato sódico (Merck).

Gelatina (Difco).

Glicerina (Merck).

Hidracina (Fluka).

Hidróxido sódico (Merck).

8-Hidroxiquinoleína (Serva).



IPTG (Bethesda Research Laboratories).

Isopropanol (Merck).

Lisozima de clara de huevo (Sigma).

2-Mercaptoetanol (Merck o Serva).

Nucleasa Bal 31 de *Alteromonas espejani* (New England Biolabs o Böhringer Mannheim).

Pantalla intensificadora de tungstato de calcio fosforado (Ilford).

Papel 3MM (Whatman).

Papel de nitrocelulosa tipo HAWP304F0 (Millipore Ibérica).

Película MAFE RP-X1 AFW50 (Manufacturas fotográficas españolas).

Película Polaroid Tipo 665 (Polaroid).

Persulfato amónico (Merck).

Piperidina (Fluka).

Piridina (Merck).

Polinucléotido kinasa del fago T4 (New England Biolabs o Böhringer Mannheim).

PVP (Sigma).

Resina de intercambio iónico AG501-X8 (D) (BioRad).

Revelador DX-90 (Kodak).

RNAse A pancreática (Worthington Biochemical).

RNAse T1 de *Aspergillus oryzae* (Böhringer Mannheim).

Sacarosa (Merck).

Sarcosil (Sigma).

SDS (Sigma).

Sephadex G50-80 (Sigma).

Sulfato de dextrano (Sigma).  
TEMED (Sigma).  
Transferasa terminal de timo de ternera (Amersham).  
Tris ácido (Sigma).  
Tris base (Sigma).  
Triton X-100 (Sigma).  
tRNA de levadura (Böhringer Mannheim).  
Urea (Sigma).  
Xilencianol (BDH).  
X-gal (Sigma).

#### 4. OTROS PRODUCTOS.

Fenol: suministrado por Probus, fue destilado y recogido sobre un volumen de Tris-HCl 10 mM, EDTA 1 mM pH 8.0 (TE), añadiéndosele 8-hidroxiquinoleína hasta una concentración final del 0.1%.

Fenol-Sevag: se mezclaron fenol saturado de TE, cloroformo y alcohol isoamílico en las proporciones 25: 24:1 (v/v).

Formamida: suministrada por Merck, se desionizó por agitación suave durante 20 min en presencia de la resina de intercambio iónico AG501-X8 (D).

#### 5. EQUIPO.

Ultracentrífuga Sorvall, modelo OTD-2.

Centrífugas Sorvall, modelo RC-5 y GLC-2.

Espectrofotómetro Shimadzu modelo UV-120-02.

Incubador de tubos selecta, modelo TemBloc.

Baños de agua con agitación giratoria de New Brunswick Co., Inc., modelo G76.

Balanzas analíticas Mettler, modelo H-15 y Sartorius tipo 2444.

pHmetro Radiometer, modelo 25 con escala expandida.

Fuentes de alimentación LKB, modelo 2103 y BioRad, modelo 160/1.6.

Estufas Heraeus.

Cámara Polaroid, modelo MP-4.

Microfugas Beckman, modelo B y Hettich, modelo Mikroliter.

Mini-monitor g.m. meter Mini-instruments, modelo 5.10.

Refractómetro Zeiss.

Secador de geles Hoefer Scientific Instruments, modelo SE 1150.

Baños termostatzados Lauda modelo RM3 y LKB modelo 2209.

Transiluminador de UV Ultraviolet Products, modelo TM36.

Miniordenador Digital pdp 11/45.

Microordenador Digital Microvax.

Microordenador Hewlett-Packard HP-85.

## MÉTODOS

## 1. PURIFICACION RAPIDA DE DNA DE PLASMIDO.

Con objeto de purificar DNA de plásmidos recombinantes de manera rápida se siguió el método descrito por Birnboim y Doly (119) con algunas modificaciones.

Un cultivo bacteriano en 15 ml de medio L con 150 ug/ml de ampicilina se creció hasta saturación durante 16 hr y se centrifugó a 8000 rpm/10 min/4 °C (HB-4). El sedimento de células se resuspendió en 1.5 ml de Tris-HCl 25 mM pH 8.0, EDTA 10 mM y sacarosa 25 mM (Tris-sacarosa) y se añadió lisozima sólida hasta una concentración final de aproximadamente 2 mg/ml, incubándose a ta unos 15 min y a continuación se enfrió 5 min en hielo. Las bacterias se lisaron con 3 ml de una disolución recién preparada de NaOH 0.2 N y SDS 1% a 4 °C durante 5 min. Se añadieron 2.25 ml de acetato potásico 3 M pH 4.8 a 4 °C y se mantuvo en hielo durante 15 min más. El precipitado se eliminó centrifugando a 12000 rpm/30 min/4 °C (HB-4) y el DNA del plásmido se precipitó del sobrenadante con 3 ml de isopropanol durante 15 min a ta. El precipitado se recogió centrifugándolo a 12000 rpm/15 min/18 °C (HB-4) y se resuspendió en 0.3 ml de Tris-HCl 25 mM pH 7.5, EDTA 1 mM, NaCl 50 mM para, a continuación, digerir el RNA contaminante con 50 ug de RNAsa A (10 mg/ml, calentada a 80 °C/10 min para inactivar las DNAasas) y 50 u de RNAsa T1 durante 1 hr a 37 °C. A continuación se extrajo dos veces con un volumen de fenol-Sevag y una vez con un volumen de cloro-

formo. La copurificación de una exonucleasa con el DNA plasmídico hace imprescindibles estos pasos de extracción para evitar la degradación del DNA cuando se digiera con enzimas de restricción.

Con el fin de eliminar los oligorribonucleótidos generados por la digestión con RNAsas, la fase acuosa de la extracción con cloroformo se cromatografió en una minicolumna centrifugada de Sephadex G50-80 de 2 ml de lecho (ver III.8.). El volumen eluído se precipitó con etanol (1/10 de volumen de acetato sódico 3 M, 3 volúmenes de etanol, -70 °C/15 min) y se centrifugó en minifuga 15min/4 °C. El sedimento se lavó con etanol, se desecó a vacío y se resuspendió en 50 ul de 0.1xTE (1xTE es Tris-HCl 10 mM pH 8.0, EDTA 1mM). Las concentraciones de DNA obtenidas, determinadas por medidas de A<sub>260</sub>, oscilaron entre 1-4 mg/ml para plásmidos de la serie pUC.

## 2. PURIFICACION DE DNA DE PLASMIDO.

Se utilizó el método del lisado claro(120). Un cultivo de 1 l crecido hasta saturación se centrifugó en un rotor GS3 a 6000 rpm/10min/4 °C y el sedimento de células se resuspendió en 40 ml de TE. Tras una centrifugación a 6000 rpm/5min/4 °C (HB-4), el sedimento se resuspendió en 5 ml de Tris-sacarosa. Se añadieron 2 ml de lisozima 5 mg/ml en EDTA 0.25 M, pH 8.0, se mezcló y se mantuvo la mezcla durante 10 min a 4 °C. Posteriormente se añadieron 8 ml de solución de

lisis (Tris-HCl 50mM, EDTA 62.5 mM, Triton X-100 0.1% pH 8.0) y se mantuvo la mezcla 15 min a 4 °C, tras lo cual se centrifugó a 16000 rpm/1 hr/4 °C (SS34).

Al sobrenadante se le añadieron 0.5 ml de EtBr (10 mg/ml) se ajustó el volumen a 16 ml con TE y se le añadieron 14.4 g. de CsCl. Tras centrifugación a 45000 rpm/40 hr/22 °C (vTi50), la banda correspondiente al plásmido se visualizó con un transiluminador de luz ultravioleta y se recogió con una aguja sobre un tubo.

El EtBr presente en la solución de DNA plasmídico se eliminó de la misma por medio de varias extracciones con igual volumen de 1-butanol saturado con TE saturado con CsCl hasta la desaparición total de color. A la fase acuosa se le añadieron posteriormente 2 volúmenes de etanol y se mantuvo la mezcla durante 4-16 hr a 4 °C.

El precipitado se recogió centrifugándolo a 10000 rpm/15 min/4 °C (HB-4), se lavó con etanol, desecó y resuspendió en 0.5 ml de 0.1xTE. Tras otra precipitación con etanol realizada de forma análoga, el DNA se resuspendió en 0.5 ml de 0.1xTE y se determinó su concentración por medida de A<sub>260</sub>. Las concentraciones de DNA obtenidas oscilaron entre 4-6 mg/ml.

### 3. DIGESTION DE DNA CON ENZIMAS DE RESTRICCION.

La digestión de DNA con enzimas de restricción se llevó a cabo según las condiciones especificadas por el fabricante

(tampón de incubación, temperatura), siendo la relación unidades de enzima/ ug de DNA a digerir generalmente mayor de 2 y el tiempo de incubación entre 2-3 hr para asegurar la digestión total. El grado de digestión de DNA se siguió en ocasiones por electroforesis en agarosa.

En el caso de digestiones con dos o más enzimas, si las condiciones de incubación eran compatibles se añadían conjuntamente. En caso contrario, tras la primera digestión y previo al tratamiento con el siguiente enzima, el DNA se precipitó con etanol.

#### 4. ELECTROFORESIS DE DNA EN GELES DE AGAROSA.

Se utilizaron geles de agarosa tipo II de dimensiones 20x20x0.5 cm o bien 10x6.5x0.8 cm. La concentración de agarosa fue 0.5% o 1% (p/v). El tampón utilizado, tanto en el gel como en la solución de electroforesis, fue TAE (Tris-acetato 40 mM pH 7.8, EDTA 2.5 mM) que contenía EtBr a una concentración de 0.5 mg/l.

La preparación de los geles se llevó a cabo según el método descrito por Maniatis (117). Las muestras, a las que se les añadió 1/10 de su volumen de tampón de aplicación (glicerol 70% (v/v), sarcosil 5% (p/v), BPB 2% (p/v), XC 2% (p/v), EDTA, 0.1 M pH 8.0), se aplicaron en los pocillos correspondientes. Las condiciones de electroforesis fueron 1.5 V/cm durante 16 hr para los geles de mayor tamaño o 8 V/cm durante 1 hr para los de menor tamaño. Para visualizar las



muestras se utilizó un transiluminador de luz ultravioleta. La fotografía del gel se llevó a cabo con una cámara Polaroid equipada con filtro rojo oscuro y naranja y película Polaroid 665.

#### 5. TRANSFERENCIA DE DNA DE GELES DE AGAROSA A PAPEL DE NITROCELULOSA.

Se siguió el método descrito por Southern (121) con ligeras modificaciones. Tras la electroforesis, el gel se sumergió en HCl 0.25 M durante 20 min. Este tratamiento produce una depurinación parcial del DNA, lo que, unido al tratamiento alcalino posterior, hace que el DNA presente en el gel se fragmente homogéneamente, y la transferencia sea independiente del tamaño. El paso siguiente fue la desnaturación del DNA por inmersión en NaOH 0.5 M, NaCl 0.6 M durante 45 min con agitación suave, tras lo que se neutralizó el gel con Tris-HCl 1 M pH 7.5 durante 1 hr. Una vez neutralizado, el gel se colocó sobre papel 3MM previamente empapado en 20xSSC (1xSSC=NaCl 150 mM, citrato sódico 15 mM) y con los bordes sumergidos en 20xSSC. Sobre el gel se puso papel de nitrocelulosa empapado en 20xSSC y encima de este un taco de unos 5 cm de espesor de papeles de filtro seco. Sobre este conjunto se colocó finalmente un vidrio con un ligero peso. La transferencia se llevó a cabo durante 16 hr a t.a. Los filtros se lavaron con 2xSSC y se secaron en estufa de vacío a 80 °C durante 2 hr.

## 6. EXTRACCION DE DNA DE GELES DE AGAROSA DE BAJO PUNTO DE FUSION.

La electroforesis en agarosa de bajo punto de fusión se llevó a cabo en geles de dimensiones 20x20x0.8 ó 1.6x6.5x0.8 cm. La concentración de la agarosa fue del 1% (p/v) y el tampón utilizado, tanto en el gel como en la solución de electroforesis, fue TAE conteniendo EtBr 0.5 ug/ml. Tras la aplicación de las muestras y su electroforesis llevadas a cabo tal como se ha descrito anteriormente (III.4), estas se visualizaron con luz ultravioleta y se cortó el trozo de agarosa que contenía el DNA a extraer.

La extracción de DNA del gel se llevó a cabo según el método descrito por Langridge et al. (122), modificado por A. González.

Soluciones H<sub>2</sub>O-CTAB y butanol-CTAB. Se prepararon de la siguiente forma: a) se equilibraron 125 ml de butanol y 125 ml de agua mezclándolos repetidas veces; b) se disolvió 1 g de CTAB en 100 ml del butanol saturado con agua; c) se añadieron a esta disolución 100 ml de agua saturada con butanol, se mezclaron ambas fases y se dejaron reposar durante 16 hr a ta; d) se separó cada fase, siendo la fase superior la solución butanol-CTAB y la inferior la solución H<sub>2</sub>O-CTAB, y se mantuvieron ambas a 37 °C.

El trozo de agarosa se fundió a 68 °C y se equilibró a 37 °C, tras lo que se le añadieron 1 volumen de H<sub>2</sub>O-CTAB y 1

volumen de butanol-CTAB. Se mezclaron con agitación fuerte (Vortex) y se centrifugaron en minifuga 1 min a 37 °C. La fase orgánica se extrajo a la vez con 1/4 de su volumen de acetato sódico 0.36 M. La fase acuosa se extrajo con 1 volumen de  $\text{Cl}_3\text{CH}$  a 4 °C, tras lo cual se precipitó con etanol el DNA presente en la misma.

#### 7. MARCAJE DE DNA CON $^{32}\text{P}$ "IN VITRO" POR DESPLAZAMIENTO DE CORTES.

El método se basa en el descrito por Rigby et al. (123) modificado posteriormente por Roop et al. (124,125).

Las reacciones, en un volumen final de 10 ul, contenían: Tris-HCl 50 mM pH 7.5,  $\text{MgCl}_2$  5 mM, 2ME 10 mM, dGTP, dTTP, dCTP 160 uM cada uno, 5'- $\alpha$ - $^{32}\text{P}$ -dATP (410 Ci/mmol) 9 uM y 0.1-0.5 ug de DNA. A la mezcla, efectuada a 4 °C, se le añadió 1 ul de DNAasa I (50 ng/ml, disuelta en Tris-HCl 10 mM, pH 7.5,  $\text{Mg}_2\text{Cl}$  10 mM y gelatina 100 ug/ml). Después de 1 min a la vez se añadieron 5 u de DNA polimerasa I y la mezcla se incubó a 14 °C durante 2 hr, al cabo de las cuales se añadieron 90 ul de TE y se cromatografió en minicolumna centrifugada para separar los nucleótidos no incorporados del DNA marcado.

## 8. CROMATOGRAFIA DE FILTRACION DE DNA EN MINICOLUMNA CENTRIFUGADA

Se siguió el método descrito por Maniatis et al (117) con ligeras modificaciones. En una jeringa de plástico desechable (1 o 2 ml) taponada con gasa se preparó una columna de Sephadex G50-80 equilibrado en TE.

La columna se centrifugó en un tubo a 2000 rpm/1 min/ta en centrífuga de mesa, para conseguir un volumen de lecho seco de 6 a 9 veces el volumen de muestra a cromatografiar. La columna se lavó con TE por centrifugación y a continuación se aplicó un volumen de TE análogo al de muestra, centrifugándose a continuación en las condiciones ya descritas. Finalmente se aplicó la muestra y se centrifugó de igual forma. El DNA eluido de la columna se usó directamente o se concentró por precipitación con etanol, según los casos.

## 9. HIBRIDACION DE DNA UNIDO A NITROCELULOSA.

El DNA unido a filtros de nitrocelulosa se hibridó con DNA radioactivo siguiendo el procedimiento descrito por Wahl y Stein (126) con algunas variaciones.

Se usaron dos procedimientos alternativos. En el primero los filtros, introducidos en una bolsa de plástico, fueron sumergidos en tampón de prehibridación: 5xSSC, 2 x solución de Denhardt (solución de Denhardt = Ficoll 0.2 g/l,

PVP 0.2 g/l, BSA 0.2 g/l). Una vez sellada la bolsa se incubaron de 2-4 hr a 68 °C. Transcurrido ese tiempo se sustituyó esta solución por 5-10 ml (aproximadamente 0.05 ml/cm<sup>2</sup> de área de filtro) de solución de hibridación: 5xSSC, 2 x solución de Denhardt, 0.1 mg/ml de DNAs o de E. coli (si los filtros contenían DNA de colonias de bacterias lisadas), y la sonda de DNA radiactivo. Ambos DNAs se habían desnaturalizado previamente calentando a ebullición durante 5 min y enfriando rápidamente a 4 °C. Una vez sellada la bolsa se incubó en baño a 68 °C durante 15-20 hr con agitación suave.

Los filtros fueron lavados con 2xSSC, 15 min a ta y seguidamente a 68 °C cambiando la solución varias veces durante 2 hr Los filtros se enjugaron sin secarlos totalmente y se sometieron, a autorradiografía tras introducirlos en una bolsa de plástico sellada.

En el segundo procedimiento la solución de prehibridación fue: 5xSSC, 5 x solución de Denhardt, formamida 50% (v/v), SDS 0.1% (p/v), fosfato sódico 50 mM, pH 6.5. La prehibridación se llevó a cabo a 42 °C durante 2-4 hr. La solución de hibridación contenía: 5xSSC, 5 x solución de Denhardt, formamida 50% (v/v), SDS 0.1 % (p/v), fosfato sódico 20 mM pH 6.5, 0.1 mg/ml de DNAs o de E. coli y la sonda radioactiva. La hibridación se realizó a 42 °C durante 15-20 hr.

Los filtros fueron lavados con 2xSSC, SDS 0.1% (p/v), cambiando la solución varias veces durante 20 min a ta. Seguidamente se lavaron con 0.1xSSC, SDS 0.1% (p/v) durante 30

min a 55 °C. Finalmente, los filtros se sometieron a autorradiografía como en el anterior apartado.

## 10. ELECTROFORESIS DE DNA EN GELES DE POLIACRILAMIDA.

### 10.1. Gel neutro.

Se utilizó este tipo de gel para separar por tamaño fragmentos de restricción marcados terminalmente con  $^{32}\text{P}$ , con fines tanto analíticos como preparativos.

Las dimensiones de los geles utilizados fueron 320x180x1.5 mm. Se prepararon según métodos descritos (117) polimerizando una solución que contenía una mezcla de acrilamida: bisacrilamida (29:1) y 0.5xTBE (1xTBE=Tris-borato 90 mM pH 8.3, EDTA 2 mM) en presencia de persulfato amónico y TEMED. Una vez polimerizado el gel se llevó a cabo una precorrida durante 30-60 min a 300 V.

A las muestras se les adicionó 1/10 de su volumen de tampón de aplicación (Tris-borato 500 mM pH 8.3, glicerina 70% (v/v), BPB 0.5%, XC 0.5%, EDTA 10 mM), y se llevó a cabo la electroforesis en tampón 0.5xTBE a 300 V.

La electroforesis se detuvo cuando la posición de los marcadores coloreados indicaba que se había obtenido una separación óptima de los fragmentos de interés, teniendo en cuenta sus tamaños y la concentración de acrilamida utilizada.

### 10.2. Gel de separación de cadenas.

Se utilizaron geles de dimensiones 160x200x1.5 mm preparados según el método descrito por Maxam y Gilbert (72). Un sedimento de DNA desecado procedente de una precipitación con etanol se resuspendió en 40 ul de tampón de aplicación: DMSO 30% (v/v), BPB 0.05% (p/v), XC 0.05% (p/v), EDTA 1 mM. Tras calentar la muestra a 90 °C/2 min, se enfrió rápidamente a 4 °C.

La muestra se cargó en un gel de acrilamida:bisacrilamida (50:1) al 5% en 0.5xTBE precorrido durante 30-60 min a 120 V. La electroforesis se llevó a cabo en 0.5xTBE a 8 V/cm o menos, cuidando de no elevar la temperatura del gel, y se detuvo cuando se estimó empíricamente que se había producido una separación óptima de los fragmentos.

El DNA presente en las muestras se desnaturalizó con el fin de que en la electroforesis ambas cadenas migrasen independientemente. Sin embargo, la electroforesis se llevó a cabo en condiciones no desnaturalizantes (gel, temperatura) para que la separación electroforética no se basase en el tamaño molecular, sino en características diferenciales de las dos cadenas (composición, estructura secundaria).

### 10.3. Gel desnaturalizante(secuenciación).

Se utilizaron geles de dimensiones 370x280x0.4 mm preparados según Maxam y Gilbert (72). Los geles contenían acrilamida:bisacrilamida (19:1), generalmente al 6%, urea 45% (p/v) y 0.75xTBE.

Las muestras se resuspendieron en 9-12  $\mu$ l de tampón de aplicación: formamida 80% (v/v), NaOH 10 mM, BPB 0.1% (p/v), XC 0.1% (p/v), EDTA 1 mM. Tras la resuspensión las muestras se calentaron a 90 °C/1 min y se enfriaron rápidamente a 4 °C.

El gel se precorrió a 1000 V durante 20-30 min, tras lo cual se cargó entre 1/3-1/4 del volumen de muestra. La electroforesis se inició a 2000 V durante aproximadamente 5 min y se continuó a 40 W hasta que el colorante XC hubo migrado hasta el final del gel. En este punto se realizó una segunda carga de otro tercio de volumen de muestra, se reinició y continuó la electroforesis como anteriormente hasta que el XC hubo recorrido 2/3 de la longitud del gel. Se realizó análogamente una tercera y última carga, y se terminó la electroforesis cuando el colorante BPB hubo recorrido aproximadamente 4/5 del gel.

La ventaja de este método de electroforesis estriba en que, utilizando un único gel de dimensiones muy manejables, las lecturas solapantes de las muestras sometidas a electroforesis durante distintos tiempos permiten determinar la secuencia de 250-400 pb a partir del extremo marcado.

En algunos casos este procedimiento se modificó (concentración del gel, condiciones de electroforesis) para aumentar la resolución en determinados intervalos de tamaños de los fragmentos.

Los geles, una vez terminada la electroforesis, se secaron a vacío a 80 °C/1 hr y se sometieron a autorradiografía.



## 11. EXTRACCION DE DNA DE GELES DE POLIACRILAMIDA.

Se siguió el método descrito por Maxam y Gilbert (72). Una vez detectada la posición de los fragmentos de interés por autorradiografía, se cortó la porción de acrilamida que los contenía, se redujo esta a una pasta y se incubó a 37 °C/16 hr con agitación vigorosa con 0.8 ml de acetato amónico 500 mM, acetato magnésico 10 mM, SDS 0.1% (p/v), tRNA de levadura 10 ug/ml.

Tras la incubación se recuperó el líquido centrifugando la mezcla, trasvasada a una punta de polipropileno desechable de 1 ml taponada con gasa, en centrífuga de mesa a 300 rpm/5 min. Se añadieron a la punta otros 0.2 ml. de la misma solución y se centrifugó en las mismas condiciones. Se mezclaron los volúmenes eluidos en ambas centrifugaciones y se les añadieron 3 ml de etanol. Tras incubar a -20 °C/16 hr se centrifugó la mezcla a 10000 rpm/15 min/4 °C(HB-4). El sedimento se lavó con etanol y se desecó, resuspendiéndose en un pequeño volumen de H<sub>2</sub>O (30-50 ul).

## 12. GENERACION DE SUBCLONES.

### 12.1. Utilizando endonucleasas de restricción.

Los insertos se obtuvieron por digestión de DNAs de clones ya disponibles con los enzimas de restricción adecuados. La cantidad de DNA utilizada fue tal que rindiera entre

0.1-0.5 ug del fragmento a subclonar. Tras la digestión (simple o múltiple), los productos de la misma se separaron por electroforesis en gel de agarosa de bajo punto de fusión y los fragmentos de interés se extrajeron según se describe en III. 6.

En el caso de que los extremos no fuesen idénticos ni compatibles con ninguno de los generados por enzimas de restricción en los sitios de clonaje múltiple de los vectores pUC, se procedió a su rellenado con DNA polimerasa I-fragmento "Klenow".

La reacción se llevó a cabo en Tris-HCl 50 mM, pH 7.4, MgCl<sub>2</sub> 10 mM, DTT 1 mM, BSA 50 mg/l, dATP, dGTP, dCTP y dTTP 75 mM cada uno en un volumen total de 50 ul, en presencia de 3 u de enzima durante 30 min a ta. Tras una extracción con fenol SEVAG y cromatografía en minicolumna centrifugada y precipitación con etanol, el DNA del inserto se resuspendió en un pequeño volumen de H<sub>2</sub>O (7-20 ul).

Los vectores se prepararon por digestión simple o múltiple, acompañadas generalmente de desfosforilación (particularmente en caso de digestión simple) simultánea a la digestión (añadiendo 2 u de fosfatasa alcalina). Los enzimas se eliminaron por extracción con fenol SEVAG y cloroformo. La cantidad de vector empleada en las mezclas de ligamiento osciló entre 10 y 20 ng.

## 12.2.Utilizando la nucleasa Bal 31.

La actividad exonucleasa de cadena doble del enzima Bal 31 se aprovechó para la generación, a partir de un fragmento ya clonado, de subclones orientados con un extremo común y el otro sucesivamente más corto, a partir de los cuales se puede obtener ordenadamente la secuencia del fragmento original, ya que las últimas bases secuenciadas en un clon solapan con las primeras determinadas en el siguiente.

En la figura 5, parte A, se resumen los distintos pasos del procedimiento: I) primer corte con una endonucleasa de restricción (A), en un extremo del fragmento total a secuenciar, en un sitio único, de forma que el plásmido resulte linearizado; II) digestión controlada con Bal 31; III) rellenado de extremos, segundo corte de restricción único en el otro extremo del fragmento (B), selección de insertos y clonaje orientado en un vector de la serie pUC. En la parte B se esquematiza el resultado: lecturas solapantes a partir de los distintos subclones delecionados.

La cantidad de DNA del clon original osciló entre 15-20 ug. La cantidad de enzima Bal 31 necesaria para generar fragmentos de tamaños adecuados en un tiempo de 15-30 min se calculó utilizando la expresión

$$dM_t/dt = -99UL/C$$

que se deriva de la ecuación de Michaelis-Menten en el caso de que el DNA a digerir tenga un tamaño inferior a 10 kpb, la concentración de DNA sea superior a 20 ug/ml y la temperatura de incubación sea 30 °C (117).

La velocidad a la que el enzima elimina nucleótidos de los extremos del DNA,  $dM_t/dt$ , se expresa en pb/min, extremo de DNA. L es la longitud del DNA en kpb, C es la cantidad de DNA a digerir en ug, y U la cantidad de enzima en unidades (117).

Si, por ejemplo, se deseaban digerir 15 ug de un fragmento de DNA de 4 kpb de longitud en 15 min para eliminar 500 pb/extremo de DNA, la cantidad necesaria de enzima sería:

$$U = 15/4 \times 500/15 \times 1/99 = 1.2 \text{ u}$$

Las equivalencias entre las unidades de enzima definidas por cada suministrador comercial se calcularon introduciendo en la expresión las condiciones de la definición de actividad. Sin embargo, la actividad enzimática real depende de factores adicionales (concentración de NaCl, tiempo de incubación, cantidad de enzima y composición del DNA).

La cantidad de DNA a digerir osciló entre 15-20 ug y la temperatura de incubación fué 30 °C. La solución de incubación tenía la siguiente composición: Tris-HCl 20 mM pH 7.2, MgCl<sub>2</sub> 12.5 mM, CaCl<sub>2</sub> 12.5 mM, NaCl 200 mM, EDTA 1 mM. La cantidad de enzima utilizada fue la teóricamente necesaria para obtener el grado de digestión deseado en un tiempo pre-determinado. Sin embargo, el tiempo real de incubación se aumentó en un 20% como margen de seguridad. Se tomaron alícuotas a distintos tiempos (incluida una antes de añadir el enzima), se inactivaron con fenol-Sevag, se extrajeron con cloroformo y se precipitaron con etanol. El rellenado de los

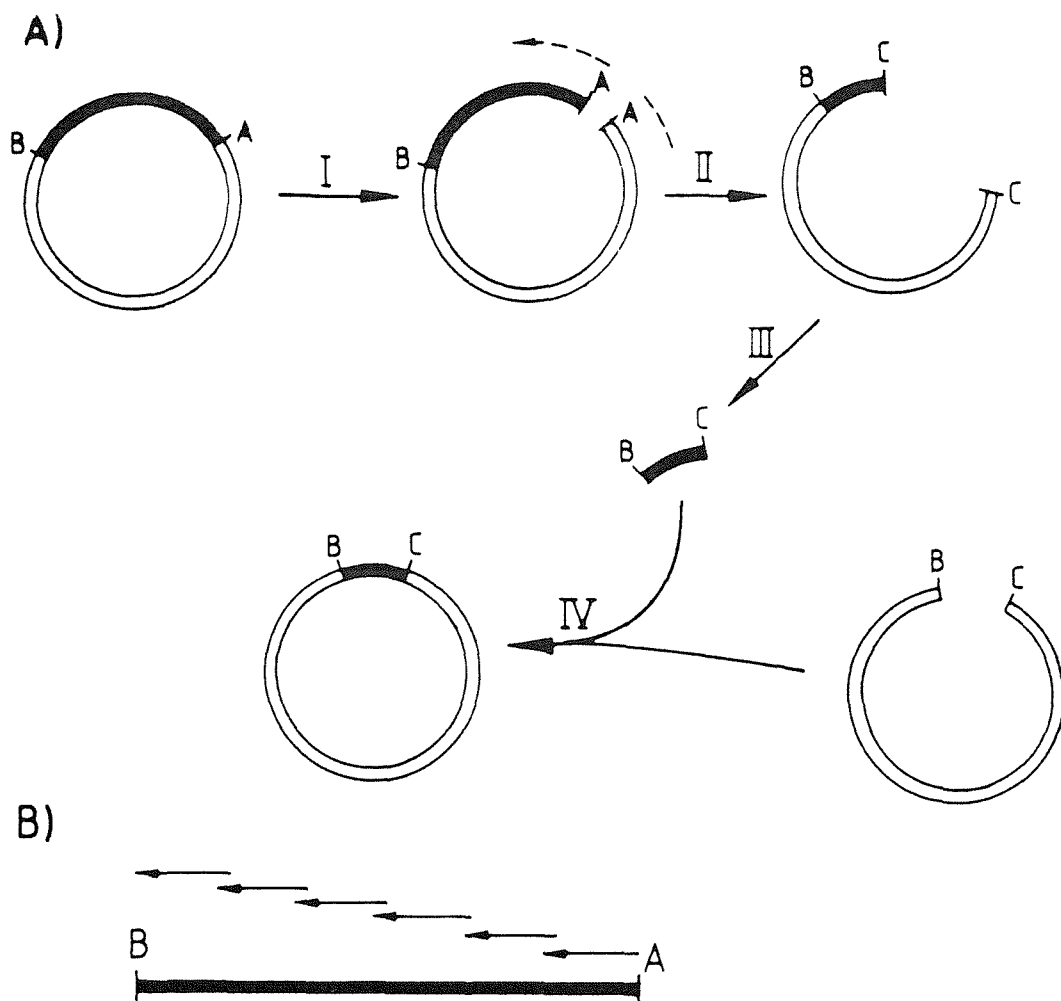


Figura 5. Estrategia de generación de subclones delecionados y orientados utilizando la nucleasa Bal 31.

A) Descripción del método. I, generación de un corte único con el enzima de restricción A. II, digestión controlada con Bal31. III, rellenado de extremos y segundo corte con el enzima B. IV, ligamiento del inserto delecionado a un vector apropiado.

B) Resultado de la estrategia: lecturas solapantes, indicadas por flechas que rinden la secuencia global del fragmento BA (112).

extremos generados por la nucleasa se llevó a cabo como se ha descrito anteriormente (III.12.1.).

Tras la segunda digestión con una enzima de restricción se separaron los vectores delecionados de los insertos delecionados y se seleccionaron los insertos delecionados adecuados por electroforesis en agarosa de bajo punto de fusión. La electroforesis se llevó a cabo en geles de 20x20x0.8 cm, según métodos ya descritos (III. 4 y III.6.). Se cortaron los trozos de agarosa que contenían insertos delecionados a intervalos aproximados de 250 pb utilizando como referencia un canal en el que se habían sometido a electroforesis en paralelo marcadores de peso molecular. La extracción del DNA de los fragmentos de agarosa y la preparación de los vectores pUC se llevó a cabo según se describió en el apartado anterior (III.12.1.).

### 13. LIGAMIENTO DE FRAGMENTOS DE DNA .

Los fragmentos de DNA con extremos compatibles fueron ligados con DNA ligasa del fago T4. La reacción se llevó a cabo en un volumen total que osciló entre 10-50  $\mu$ l y que contenía: Tris-HCl 20 mM pH 7.6, MgCl<sub>2</sub> 10 mM, DTT 10 mM, ATP 0.6 mM, los fragmentos de DNA a ligar y 1 u de DNA ligasa. El tiempo de incubación fue variable (2-16 hr) y la temperatura de incubación 16 °C.

#### 14. TRANSFORMACION DE E. COLI CON PLASMIDOS.

Para la transformación tanto con plásmidos como con formas replicativas de DNAs de fagos (M13) se siguió el método descrito por Mandel y Higa (127) con ligeras modificaciones.

La preparación de células competentes congeladas para la transformación con plásmidos se realizó siguiendo el método descrito por Morrison (128).

Un cultivo de 1 l de la bacteria correspondiente crecido hasta alcanzar una  $DO_{550}=0.5$  (aproximadamente  $5 \times 10^8$  bac/ml), se enfrió rápidamente a 4 °C y se centrifugó a 8000 rpm/8 min/4 °C (GSA). El sedimento se resuspendió suavemente en 250 ml de  $MgCl_2$  0.1 M a 4 °C. Las bacterias resuspendidas se centrifugaron en las mismas condiciones y el sedimento se resuspendió suavemente en 250 ml de  $CaCl_2$  0.1M a 4 °C y se mantuvo a la misma temperatura durante 20 min, tras los cuales se centrifugó en las mismas condiciones. Finalmente el sedimento de bacterias se resuspendió en 43 ml de  $CaCl_2$  0.1 M y 7 ml de glicerol. Se distribuyó en alícuotas de 0.5 ml y se congeló rápidamente a -70° C.

Para tranformar, se dejaron descongelar a 4 °C las bacterias competentes, y se añadieron 0.125 ml de suspensión de las mismas en cada mezcla de ligamiento o control de transformación. Los controles de transformación fueron habitualmente: a) transformación con DNA del vector sometido a ligamiento en las mismas condiciones que la mezcla de ligamien-

to, salvo la presencia del inserto a subclonar; b) transformación en ausencia de DNA.

Tras una incubación de 30 min a 4 °C, se calentaron las bacterias a 42 °C durante 2 min, se volvieron a enfriar a 4 °C y se les añadió 1 ml de medio L. Tras una incubación a 37 °C/45 min, las bacterias se centrifugaron en minifuga durante 30 segundos, el sedimento se resuspendió en 0.2 ml de medio L y se sembraron en placas de medio L conteniendo ampicilina (150 mg/l) y, en el caso de recombinantes construidos utilizando como vector alguno de la serie pUC, se detectó la presencia de inserto sembrando en placas de medio L-amp-X-gal.

Para transfectar bacterias JM101 con recombinantes obtenidos usando como vectores M13mp8 o mp9, se añadieron a las mezclas de ligamiento y controles 0.1 ml. de bacterias JM101 competentes y se incubaron durante 40 min a 4 °C. Tras una incubación posterior de 2 min a 42 °C, se añadieron 10 ul de IPTG 100 mM, 50 ul de X-gal (2% p/v), 0.2 ml células JM101 en crecimiento exponencial y 3 ml de agar de plaqueo. Se plaquéo la mezcla rápidamente sobre de medio L. Una vez enfriado el agar de plaqueo, las placas se incubaron durante 16 hr a 37 °C.



## 15. DETECCION E IDENTIFICACION DE SUBCLONES.

### 15.1. Clonados en M13mp8 y 9.

Se incubaron cultivos de 2 ml de bacterias JM101 en crecimiento exponencial con placas de lisis incoloras procedentes de la transfección con recombinantes con agitación vigorosa durante 6 hr a 37 °C. Se semipurificó la forma replicativa de los fagos recombinantes siguiendo el procedimiento ya descrito (III.1.) adaptado al menor volumen del cultivo.

La presencia de insertos del tamaño esperado se determinó por digestión con los enzimas apropiados del DNA así obtenido. Los clones positivos se crecieron en cultivos de 1 l y se obtuvo la forma replicativa según el método ya descrito (III.2.). La identificación de los insertos se realizó por digestión, electroforesis en geles de agarosa, transferencia a papel de nitrocelulosa e hibridación con el inserto global a subclonar marcado por desplazamiento de cortes.

### 15.2. Clonados en vectores de la serie pUC.

Se siguieron dos procedimientos, según el resultado obtenido en la transformación. Si la relación colonias blancas/colonias azules en la transformación con la mezcla de ligamiento fue muy superior a la observada en la transformación con control de vector religado, se seleccionaron al azar varias colonias blancas y se semipurificó DNA plasmídico. Este DNA se sometió a digestión con enzimas de restric-

ción y se analizaron los fragmentos obtenidos por electroforesis en geles de agarosa, con el fin de confirmar la presencia de insertos.

Si la comparación de ambas relaciones no hacía previsible la presencia mayoritaria de clones recombinantes, se detectaron estos siguiendo el método de Grunstein y Hogness (129).

Se obtuvieron réplicas de las placas procedentes de la transformación en filtros de nitrocelulosa colocando estos ,secos ,sobre las placas y manteniéndolos hasta que los filtros estuvieron húmedos. Las placas se recrecieron a 37 °C durante 6 hr.

Los filtros réplica conteniendo las colonias bacterianas se colocaron con éstas hacia arriba sobre papel 3MM empapado en NaOH 0.5 M durante 5-10 min con el fin de de lisar las bacterias. Posteriormente se neutralizaron pasándolos a papeles empapados en Tris-ClH 0.5 M pH 7.5 (2x2 min) y finalmente a papel empapado en Tris-ClH 0.5 M, NaCl 1.5 M pH 7.5 (5-10 min). Los filtros se secaron al aire y una vez secos se incubaron a 80 °C durante 2 hr a vacío. Los filtros así tratados se hibridaron con fragmentos de DNA marcados radiactivamente por desplazamiento de cortes. A partir de las colonias identificadas como positivas se semipurificó DNA plasmídico, analizándose este como en el primer procedimiento.

## 16. MARCAJE TERMINAL DE FRAGMENTOS DE RESTRICCIÓN CON $^{32}\text{P}$ .

### 16.1. Marcaje en 5' de extremos 5'-protuberantes con polinucleótido-kinasa.

Tras la digestión del DNA (generalmente 15-20 ug de DNA de plásmido) con el enzima de restricción apropiado, se añadió TE hasta un volumen total de 100 ul y se incubó la mezcla a 37 °C/1 hr en presencia de 2 u de fosfatasa alcalina.

La inactivación de los enzimas se llevó a cabo incubando la mezcla a 68 °C/20 min en presencia de ácido nitriloacético 10 mM pH 7.6, SDS 0.5% (p/v), extrayendo dos veces con fenol-Sevag y una con  $\text{Cl}_3\text{CH}$ , y se cromatografió la fase acuosa en minicolumna centrifugada. El DNA presente en el eluido se precipitó con etanol, y se lavó y desecó a vacío. El sedimento de DNA se resuspendió en 10 ul de  $\text{H}_2\text{O}$ , a los que se añadieron 2 ul de 10 x tampón de kinasa (Tris-HCl 500 mM,  $\text{MgCl}_2$  100 mM, DTT 50 mM pH 7.5), 7 ul de 5'- $^{32}\text{P}$ -ATP (3000 Ci/mmol, 10 mCi/ml) y 1 ul de polinucleótido-kinasa (5 u). La reacción se incubó a 37 °C durante 45 min, transcurridos los cuales se separó el DNA de la radiactividad no incorporada por cromatografía en minicolumna centrifugada. El DNA eluido se concentró por precipitación con etanol.

### 16.2. Marcaje en 3' de extremos 5'-protuberantes con el fragmento "Klenow" de la DNA polimerasa-I de E. coli.

Tras la digestión del DNA con el enzima de restricción adecuado (generalmente 5-20 ug de DNA de plásmido), el enzi-

ma se eliminó por una extracción con fenol-Sevag, otra con  $\text{Cl}_3\text{CH}$  y el DNA se concentró por precipitación con etanol. Al sedimento se le añadieron 20  $\mu\text{l}$  de  $3'\text{-}\alpha\text{-}^{32}\text{P}\text{-dNTP}$  (410 Ci/mmol, 1 mCi/ml) y se desecó a vacío. La reacción de marcaje se llevó a cabo en 20  $\mu\text{l}$  de Tris-HCl 50 mM pH 7.4,  $\text{MgCl}_2$  10 mM, DTT 1 mM y BSA 50  $\mu\text{g/ml}$  en presencia de 0.5-2 u de fragmento "Klenow" durante 3 hr a 14 °C.

Si en el extremo generado por el enzima se podía insertar más de un nucleótido marcado se incubaba, tras la adición de 1.5  $\mu\text{l}$  de dNTP 1 mM, durante 10 min más para asegurar la homogeneidad de los extremos. La reacción se detuvo por incubación durante 10 min a 70 °C y a continuación se procedió como en el caso anterior.

#### 16.3. Marcaje en 3' de extremos romos o 3'-protuberantes con transferasa terminal.

Tras la digestión del DNA (aproximadamente 20  $\mu\text{g}$  de DNA de plásmido) y eliminación del enzima de restricción como en el apartado , se resuspendió el DNA en 35  $\mu\text{l}$  de  $\text{H}_2\text{O}$ . Se añadieron 5  $\mu\text{l}$  de 10xtampón de transferasa terminal (cacodilato sódico-HCl 1.4 M pH 7.5,  $\text{CoCl}_2$  10 mM, DTT 1 mM), 5  $\mu\text{l}$  de  $5'\text{-}\alpha\text{-}^{32}\text{P}\text{-ddATP}$  (3000 mCi/mmol, 10 mCi/ml) y 5  $\mu\text{l}$  de transferasa terminal (2 u/ $\mu\text{l}$ ). Tras la incubación durante 1 hr a 37 °C se purificó el DNA marcado por cromatografía en minicolumna centrifugada y se concentró por precipitación con etanol.

## 17. OBTENCION DE FRAGMENTOS DE DNA MARCADOS CON $^{32}\text{P}$ EN UN UNICO EXTREMO.

### 17.1. Corte secundario con endonucleasas de restricción.

El DNA marcado terminalmente según los métodos descritos anteriormente se digirió con un segundo enzima de restricción para generar fragmentos con un único extremo marcado. La digestión se llevó a cabo en un volumen máximo de 50 ul. Los fragmentos obtenidos se separaron por electroforesis en agarosa de bajo punto de fusión y aquellos destinados a las reacciones de secuenciación se extrajeron del gel, añadiendo a la precipitación con etanol 1 ug de DNAs sonificado para favorecer la precipitación cuantitativa. El sedimento de DNA desecado se resuspendió en un pequeño volumen de  $\text{H}_2\text{O}$  (30-60 ul).

### 17.2. Electroforesis en gel de separación de cadenas.

El fragmento de DNA marcado en ambos extremos, aislado por electroforesis en agarosa de bajo punto de fusión, se sometió a electroforesis en un gel de separación de cadenas (III.10.2). La detección de los fragmentos de interés se realizó por autorradiografía del gel humedo y el DNA marcado se extrajo del gel de poliacrilamida por difusión (III.11).

## 18. REACCIONES DE DEGRADACION QUIMICA DE DNA (SECUENCIACION).

Las reacciones de modificación química específica de

bases en condiciones controladas y de rotura del DNA en las bases modificadas se llevaron a cabo según el método descrito por Maxam y Gilbert (72), con ligeras modificaciones.

El procedimiento seguido se esquematiza en la tabla I. El DNA marcado se repartió en cuatro tubos Eppendorf , según se describe en dicha tabla, se sometió a las reacciones de modificación durante tiempos que dependieron del tamaño del fragmento ( tiempos más prolongados para fragmentos de menor tamaño) y los reactivos se eliminaron por precipitación rápida con etanol ( G, T y C) o por liofilización ( A). La composición de las soluciones empleadas fue la siguiente:

-Incubación con DMS ( tampón DMS): cacodilato sódico 50 mM pH 8.0, EDTA 1 mM.

-Inactivación de DMS ( DMS stop): acetato sódico 1.5 M pH 7.0 , 2ME 1.0 M, tRNA de levadura 400 mg/l.

-Depurinación ácida (pir for): ácido fórmico 1 M ajustado con piridina a pH 2.0 (concentración final de ácido fórmico, 0.1 M).

-Parada de HZ ( HZ stop): acetato sódico 0.36 M pH 7.0, EDTA 0.1 mM, tRNA de levadura 75 mg/l.

El siguiente paso fue la reacción de rotura con piperidina 1 M, eliminándose luego esta por liofilización. Las muestras liofilizadas se sometieron a electroforesis en gel de secuenciación ( III.10.3) tras ser resuspendidas en tampón de muestra.

## 19. AUTORADIOGRAFIA.

Sobre los geles o los filtros de nitrocelulosa con DNA radiactivo se colocó una película MAFE AFW50 y se guardó en una caja de exposición en la oscuridad el tiempo apropiado.

Cuando el nivel de radiactividad era bajo se utilizaron pantallas amplificadoras de tungstato de calcio fosforado y la exposición se realizó a -70 °C.

Las películas se trataron con revelador DX80 durante 1-5 min, se lavaron con agua y se sumergieron en fijador FX40 durante 5 min. Después de lavarse con agua se secaron a t.a. Alternativamente se usó un aparato revelador automático.

## 20. NOMENCLATURA DE SUBCLONES.

La nomenclatura de los subclones intenta resumir sus características:

- La primera sigla denota el vector empleado (p5 = pBR325, p8 = pUC8, p9 = pUC9, y así sucesivamente).

- Si el inserto se ha obtenido utilizando enzimas de restricción, se denotan estas de forma abreviada y generalmente a continuación el tamaño del inserto en kpb.

- Si el inserto se ha obtenido utilizando la nucleasa Bal 31, la cifra o letra tras la abreviatura del vector hace referencia al inserto original y la letra mayúscula indica el tamaño del inserto delecionado progresivamente (A>B>C , etc.).

TABLA I. Reacciones de degradación química de DNA.

G	G + A	T + C	C
200 $\mu$ l tampón DMS 5 $\mu$ l $^{32}$ P DNA 1 $\mu$ l DMS	10 $\mu$ l agua 10 $\mu$ l $^{32}$ P DNA 2 $\mu$ l pir for	10 $\mu$ l agua 10 $\mu$ l $^{32}$ P DNA 30 $\mu$ l HZ	15 $\mu$ l NaCl saturado 5 $\mu$ l $^{32}$ P DNA 30 $\mu$ l HZ
20°C, $\geq$ 20 s	37°C, $\geq$ 4 min	20°C, $\geq$ 2 min	20°C, $\geq$ 2 min
50 $\mu$ l DMS stop 750 $\mu$ l EtOH Hielo seco 15' G 15' Resusp 250 $\mu$ l (0.3 M AcNa) 750 $\mu$ l EtOH 15' hielo seco G 15' Lavar pellet G 2' desecar	G 1" Congelar Liofilizar	200 $\mu$ l HZ stop 750 $\mu$ l EtOH Hielo seco 15' G 15' Resusp 250 $\mu$ l (0.3 M AcNa) 750 $\mu$ l EtOH 15' hielo seco G 15' Lavar pellet G 2' desecar	200 $\mu$ l HZ stop 750 $\mu$ l EtOH Hielo seco 15' G 15' Resusp 250 $\mu$ l (0.3 M AcNa) 750 $\mu$ l EtOH 15' hielo seco G 15' Lavar pellet G 2' desecar

- 100  $\mu$ l piperidina 1 M, 90°C, 30 min.

- Liofilizar



## 21. PROGRAMAS DE ORDENADOR.

Los siguientes programas de ordenador fueron desarrollados por M. Caballero para ser utilizados en un miniordenador Digital pdp.

ESN: permite visualizar y corregir la secuencia de nucleótidos almacenada.

LIS: permite visualizar e imprimir la secuencia almacenada.

FYC: calcula la composición de nucleótidos y las frecuencias de dinucleótidos de una determinada secuencia.

BTR: localiza y representa sobre la secuencia de nucleótidos la posición de los tripletes de iniciación y terminación en las seis fases posibles de lectura.

PES: es un programa auxiliar de predicción de estructura secundaria de proteínas. Calcula diversos parámetros conformacionales medios de una determinada secuencia de aminoácidos utilizando los parámetros conformacionales individuales calculados por Chou y Fasman a partir de proteínas con estructuras tridimensionales conocidas (130).

Los parámetros conformacionales medios son:

- Frecuencias promedio de estructura helicoidal para todos los grupos de 4, 5 y 6 residuos consecutivos ( PA4, PA5 y PA6 ).

- Frecuencias promedio de estructura en  $\beta$  para todos los grupos de 4, 5 y 6 residuos consecutivos ( PB4, PB5 y PB6 ).

- Frecuencia promedio de estructura en giro  $\beta$  para todos los grupos de cuatro residuos consecutivos (Pt).

- Probabilidad de aparición de giro  $\beta$  para todos los grupos de cuatro residuos consecutivos (pt).

El programa imprime estos parámetros conformacionales medios asignados al primer residuo de cada grupo. Imprime además:

- El tipo de residuo (favorecedor, indiferente o rompedor de estructuras  $\alpha$  o  $\beta$ ).

- Las frecuencias individuales de aparición de los residuos en estructuras  $\alpha$ ,  $\beta$  y giro  $\beta$ .

- Las frecuencias individuales de aparición de cada residuo en los límites N- y C-terminales de hélices  $\alpha$  o cadenas  $\beta$ .

- Las frecuencias individuales de aparición en las zonas adyacentes N- o C-terminales a estas dos estructuras.

Finalmente, representa gráficamente los valores PA4 y PB4 respecto a la posición en la secuencia de aminoácidos.

RSN: calcula el número de diferencias existente entre dos secuencias de nucleótidos que han sido alineadas previamente introduciendo adiciones y/o deleciones para maximizar la homología. Las diferencias se contabilizan siguiendo los criterios de Fitch y Margoliash (131): una delección de un triplete equivale a una diferencia de un nucleótido. Permite asimismo imprimir la secuencias alineadas y su traducción a aminoácidos.

RSA: es un programa análogo a RSN que efectúa los mismos cálculos para dos secuencias de aminoácidos.

ANALYSEQ: es un programa desarrollado por Staden (132) para el análisis de secuencias de nucleótidos. En este trabajo se ha utilizado para calcular las composiciones de aminoácidos y las frecuencias de uso de codones de las proteínas codificadas por las distintas fases de lectura abierta.

Los siguientes programas fueron elaborados por M. A. Vega y utilizados en un microordenador HP-85 equipado con un terminal gráfico.

DET: calcula y representa graficamente los valores medios de hidrofiliidad para todos los grupos de seis residuos consecutivos de una secuencia de aminoácidos de acuerdo con el método de Hopp y Woods (133). El valor de hidrofiliidad medio se asigna al cuarto residuo de cada grupo.

ANTIG: calcula y representa los valores de antigenicidad potencial media para todos los grupos de siete residuos consecutivos de una secuencia de aminoácidos de acuerdo con el método de Welling et al.(134). El valor medio se asigna al cuarto residuo de cada grupo.

HID: calcula y representa el índice hidropático medio de todos los grupos de nueve residuos consecutivos de una secuencia de acuerdo con el método de Kyte y Doolittle (135). El valor medio se asigna al quinto residuo de cada grupo.

SEC: es un programa de predicción de estructura secundaria de proteínas según el método direccional de Garnier et al. (136), en el que la conformación de un determinado resi-

duo  $i$  se predice teniendo en cuenta la influencia de los residuos  $i-8$  a  $i+8$  de la secuencia. Las constantes de decisión utilizadas para las conformaciones de hélice  $\alpha$  y cadena  $\beta$  dependieron en cada caso de los contenidos globales de dichas estructuras predichos por el método sin utilizar constantes de decisión.

Los siguientes programas fueron creados por H. Martínez para ser utilizados en un microordenador Microvax.

NUCSEC: calcula la estructura secundaria termodinámicamente más estable de una secuencia de nucleótidos de acuerdo con las reglas de Tinoco (137).

FASTP: es un programa de búsqueda de secuencias de aminoácidos homólogos basado en el algoritmo desarrollado por Wilbur y Lipman (138).

## 22. CALCULO DEL PUNTO ISOELCTRICO TEORICO DE UNA PROTEINA A PARTIR DE SU COMPOSICION DE AMINOACIDOS.

Se llevó a cabo utilizando los valores de  $pK_a$  de las cadenas laterales de los aminoácidos individuales (139).

## 23. REPRESENTACION EN FORMA DE DENDROGRAMA DEL GRADO DE DIVERGENCIA EXISTENTE ENTRE DISTINTAS SECUENCIAS DE NUCLEOTIDOS.

Se utilizó el método de Gibbs y Fenner (140). En el dendrograma cada punto de ramificación representa la media

de los números de diferencias existentes entre todas las secuencias conectadas a través de ese punto.

#### 24. PREDICCIÓN DE PEPTIDO SEÑAL.

Se utilizó en primer lugar el método desarrollado por McGeoch (141), que analiza el carácter hidrofóbico de la zona N-terminal de la secuencia en base a dos parámetros: longitud del segmento no cargado de mayor longitud e índice hidropático medio máximo de dicho segmento. La representación bidimensional de estos dos valores discrimina adecuadamente a la mayoría de las proteínas cuya traducción se lleva a cabo en ribosomas unidos al retículo endoplásmico.

Las características de las secuencias N-terminales predichas como péptidos señal de acuerdo con este método se compararon con las que presentan los péptidos señal de 39 proteínas analizadas por Perlman y Halvorson (142). Estas características son: presencia de un segmento hidrofóbico flanqueado por residuos cargados, composición de aminoácidos del segmento hidrofóbico y estructura secundaria predicha para los aminoácidos adyacentes en posición C-terminal al segmento hidrofóbico. Se analizó también la presencia de secuencias consenso de reconocimiento por endopeptidasas.

## 25. PREDICCIÓN DE ESTRUCTURA SECUNDARIA DE PROTEINAS.

Se llevó a cabo utilizando conjuntamente los métodos de Garnier et al.(136) y Chou y Fasman (130) con la ayuda de los programas SEC y PES.

Para la asignación de los tetrapéptidos en conformación de giro  $\beta$  se establecieron como condiciones adicionales, además de las utilizadas por Chou y Fasman:a) que las frecuencias promedio de giro  $\beta$  de los tetrapéptidos adyacentes y solapantes en tres residuos al asignado fuesen superiores a 1.00;b) que alguno de los tetrapéptidos adyacentes y solapantes hasta en un residuo al asignado presentase una frecuencia promedio superior a 1.00 y una probabilidad de aparición de giro  $\beta$  superior a  $0.75 \times 10^{-4}$ .

Aquellos tetrapéptidos así seleccionados como giros  $\beta$  localizados en zonas predichas como hélices  $\alpha$  o cadenas  $\beta$  por alguno de los dos métodos se consideraron más en detalle. Estos tetrapéptidos, o bien contenían residuos rompedores de estructuras  $\alpha$  o  $\beta$  segun Chou y Fasman, o bien las diferencias entre los potenciales de ambas estructuras según Garnier et al. no eran significativas, o presentaban ambas características. Por lo tanto se predijeron como giros  $\beta$ .

Aquellos segmentos en los que la asignación de la estructura secundaria ( $\alpha$  o  $\beta$ ) era distinta según cada método se consideraron más en detalle. Se analizaron las diferencias existentes entre los parámetros indicativos de cada estructura definidos por cada uno de los métodos y se asignó

la estructura que presentaba la mayor diferencia entre dichos parámetros.

Algunos tetrapéptidos predichos como giros  $\beta$  por el método de Chou y Fasman, pero que no cumplían los requisitos adicionales descritos anteriormente, se predijeron finalmente como giros  $\beta$  si sus residuos flanqueantes resultaban preflanqueantes resultaban predichos en conformación  $\alpha$  o  $\beta$ .

## RESULTADOS



# 1. MAPEO DE SITIOS DE RESTRICCIÓN EN LOS FRAGMENTOS TERMINALES CLONADOS (ECORI-K' Y ECORI-D').

Como se observa en los mapas de restricción del genoma del VPPA representados en la figura 1, el sitio de restricción conocido más próximo al extremo izquierdo del genoma se encuentra a una distancia de 4.8 kpb de dicho extremo.

Con el fin de obtener subclones de menor tamaño del fragmento terminal izquierdo clonado, se realizó la búsqueda de sitios de restricción en el fragmento terminal izquierdo clonado EcoRI-K'. Estos subclones serían más adecuados tanto para secuenciar a partir de los sitios de restricción detectados, como para generar a partir de dichos subclones otros delecionados y orientados que nos rindieran lecturas solapantes.

El caso del extremo derecho es similar, pues el fragmento de restricción terminal mapeado de tamaño menor es el Pvu-I(2.5 kpb).

Los enzimas utilizados fueron BamHI, HincII, HindIII, PstI, SacI y XbaI. Se realizaron digestiones sencillas con cada uno de los enzimas y dobles con cada enzima y EcoRI, de DNA de los recombinantes p5RK'1 y p5RD'2 y los fragmentos obtenidos se analizaron por electroforesis en gel de agarosa al 0.5%. La digestión doble sirvió para identificar los fragmentos generados por los enzimas que reconocieran sitios presentes en los insertos de los recombinantes, comparando

los productos de su digestión doble con los generados por la digestión doble del vector pBR325. En la figura 6 se muestran, como ejemplo, los productos obtenidos en el análisis con el enzima HindIII. El tamaño de los fragmentos se calibró tomando como referencia las movilidades de fragmentos de tamaño conocido (DNA de  $\lambda$  digerido con HindIII). Los resultados para los seis enzimas escogidos se muestran en la tabla II. En todos los casos la suma de los tamaños calibrados coincide aproximadamente con el tamaño, determinado asimismo por electroforesis, de los insertos terminales (EcoRI-K', 3.8 kpb y EcoRI-D', 9 kpb).

Los tamaños de los fragmentos resultantes de la digestión de p5RK'1 con HindIII permitieron la orientación de los sitios reconocidos por dicho enzima respecto al mapa de restricción del vector, teniendo en cuenta la posición del único sitio Hind III presente en el mismo. El mapa de restricción para los enzimas HindIII y EcoRI de p5RK'1 se muestra en la figura 6.

Para localizar la presencia de la repetición terminal invertida (RTI) en los fragmentos HindIII del fragmento clonado EcoRI-K', los productos de las digestiones sencillas HindIII y EcoRI de p5RK'1 se sometieron a electroforesis en gel de agarosa al 0.5%, se transfirieron a nitrocelulosa y se hibridó el filtro que los contenía con DNA del fago recombinante LRD'16 (que contiene el fragmento terminal EcoRI-D') marcado radiactivamente con  $^{32}\text{P}$ .

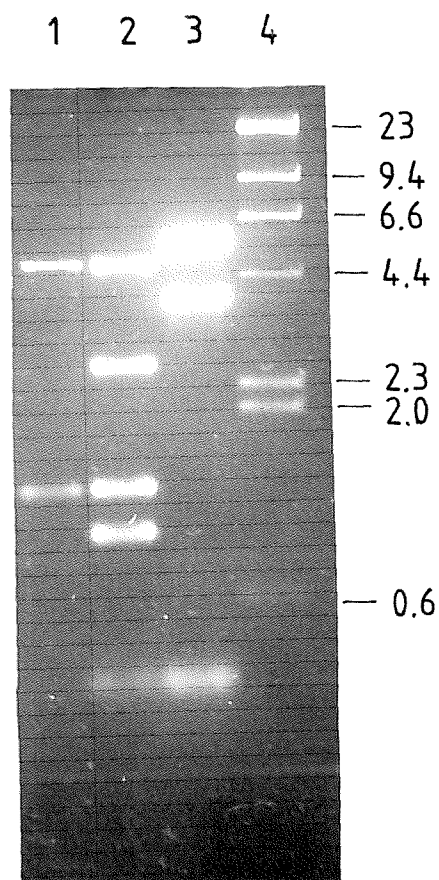
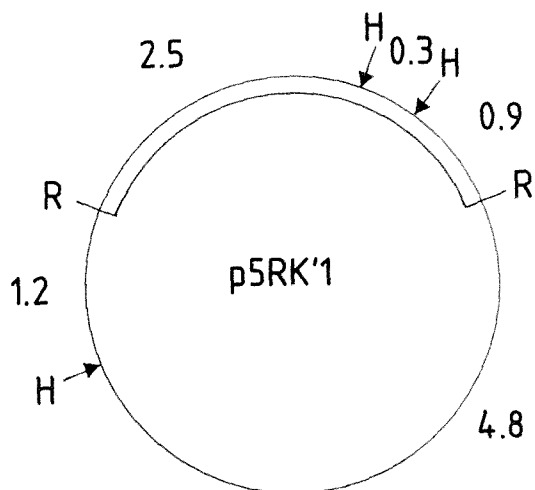
**A****B**

Figura 6. A) Análisis de los productos de digestión del recombinante p5RK'1 con los enzimas HindIII y EcoRI.

1) pBR325 digerido con HindIII y EcoRI; 2) p5RK'1 digerido con HindIII y EcoRI; 3) p5RK'1 digerido con HindIII; 4)  $\lambda$  digerido con HindIII. A la derecha se indican los tamaños de los fragmentos de  $\lambda$  en kpb.

B) Mapa de restricción del recombinante p5RK'1. Se muestran los sitios HindIII y EcoRI, así como las distancias que los separan (en kpb). El inserto de VPPA se señala con trazo más grueso.

TABLA II. Sitios de restricción en los fragemntos termina-  
les EcoRI clonados.

K'			D'	
Enzima	N°de sitios	Tamaños de los fragmentos (kpb)	N°de sitios	Tamaños de los fragmentos (kpb)
BamHI	0	2.5/0.9/0.3	1	4.5/4.3
HincII	0		3	2.9/2.8/1.7/1.2
HindIII	2		2	8.1/0.45/0.4
PstI	0		3	4.4/1.7/1.7/1.1
SacI	0		1	4.9/4.0
XbaI	0		2	3.6/2.9/2.5

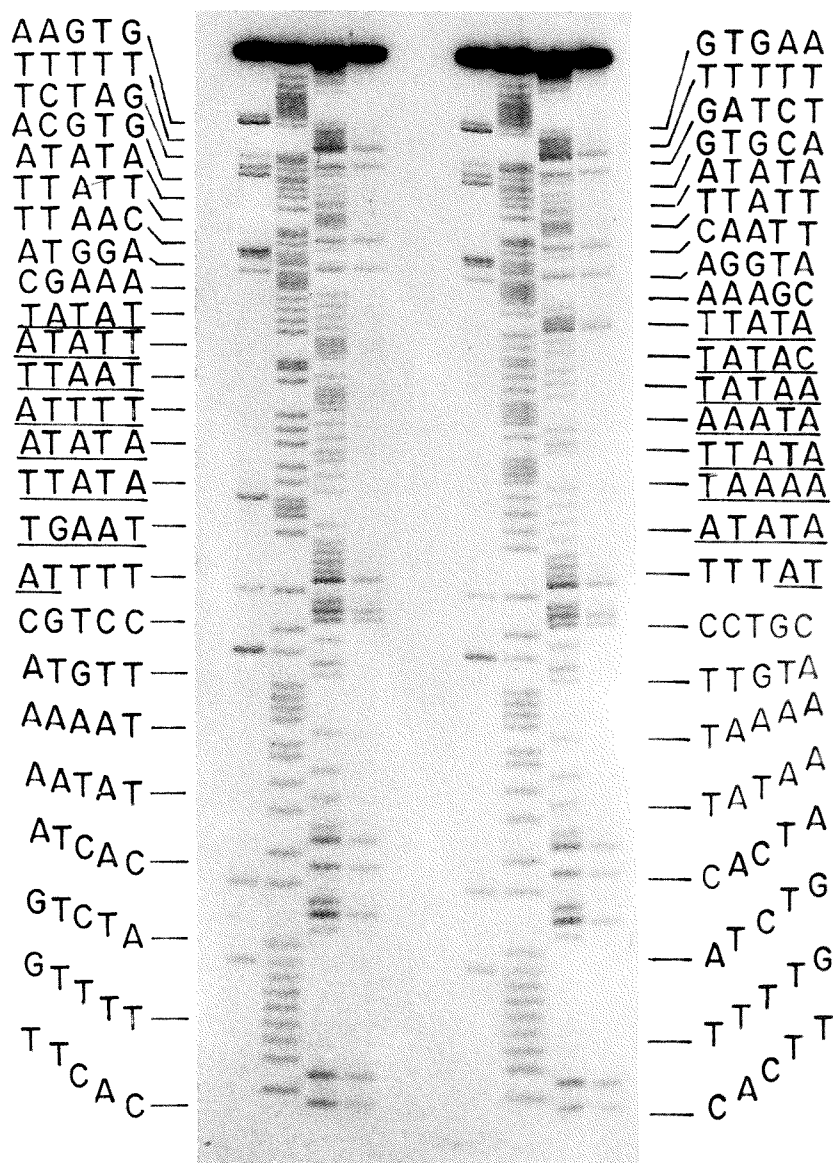


Figura 7. Orientación de los sitios HindIII en el fragmento EcoRI-K' clonado a) Fragmentos de restricción de P5RK'1 separados electroforéticamente. 1) Digestión con EcoRI. 2) Digestión con HindIII. b) Hibridación de dichos fragmentos, transferidos a nitrocelulosa, con DNA del fago recombinante LRD'16, marcado radiactivamente con  $^{32}\text{P}$  por desplazamiento de cortes. Canales 1 y 2, como en la parte a).

El resultado se muestra en la figura 7. El DNA marcado hibridó con el inserto completo (canal 1) y con el fragmento de 3.7 kpb que contiene el fragmento EcoRI-HindIII de 2.5 kpb de K' (canal 2). Dicho fragmento contiene por lo tanto la repetición terminal invertida, con lo que los sitios Hind III quedan localizados respecto al genoma del VPPA.

## 2. GENERACION DE SUBCLONES A PARTIR DE LOS FRAGMENTOS TERMINALES CLONADOS.

La generación de subclones perseguía dos tipos de fines, encaminados ambos a facilitar la secuenciación por el método de degradación química:

a) La creación de extremos marcables por los métodos descritos (III.16) en zonas de secuencia en las que no se detectaba la presencia de sitios de restricción apropiados, utilizando el método 12.2 que emplea la nucleasa Bal 31.

b) La disponibilidad de recombinantes en los que los sitios de restricción detectados pudiesen ser utilizados más cómodamente que en los clones originales, en el caso de subclones generados utilizando enzimas de restricción (III.12.2).

Los subclones conteniendo insertos de menor tamaño, con menor número de sitios de restricción y con los sitios de restricción de la zona de clonaje múltiple de los vectores pUC flanqueando el inserto facilitan la obtención de fragmentos marcados terminalmente (III.16) en un único extremo por el método del corte secundario (III.17.1) mas cómodo y efectivo que el de la separación de cadenas (III.17.2).

### 2.1.Fragmento terminal izquierdo (EcoRI-K').

El procedimiento de obtención de cada uno de los subclones generados y utilizados en el curso de la secuenciación del extremo izquierdo del genoma se muestra en la tabla IV. La tabla incluye un número identificador de cada clon en figuras subsiguientes, el nombre asignado a cada subclon, las características del inserto, la procedencia de éste, el vector en el que se insertó y en qué sitios de restricción del vector se llevó a cabo la inserción.

La figura 8 muestra la localización de los insertos de los subclones con respecto a un mapa de restricción del fragmento terminal izquierdo.

La ausencia de sitios de restricción que permitieran el subclonaje directo en vectores de la serie pUC, especialmente en la zona más próxima al extremo del genoma (el sitio HindIII terminal se encuentra a 2.5 kpb del extremo, IV.1), obligó a abordar la generación de subclones delecionados.

El clon original contenía el fragmento terminal EcoRI<sub>1</sub>-HindIII insertado en M13mp9 (M13mp9RH 2.5). El subíndice 1 indica que el sitio de restricción EcoRI<sub>1</sub> proviene de la adición de conectores ("linkers") utilizada en el clonaje de los fragmentos EcoRI-terminales (58). La diferencia de tamaño entre vector e inserto (> 4kpb) permitió la separación eficiente de insertos delecionados y vectores delecionados por electroforesis en agarosa de bajo punto de fusión (III.12.).

TABLA III. Subclones generados y utilizados para la secuenciación del fragmento terminal izquierdo.

Nº	NOMBRE	INSERTO	PROCEDENCIA	VECTOR	SITIOS DE INSERCIÓN
1	M13mp9-RH2.5	R <sub>1</sub> H 3.0	p5RK'1	M13mp9	R, H
2	p12 2.5B	Bal31/H	M13mp9 RH2.5	p12	Hc, H
3	p12 2.5C	"	"	"	"
4	p12 2.5D	"	"	"	"
5	p12 2.5E	"	"	"	"
6	p12 2.5F	"	"	"	"
7	p12 2.5G	"	"	"	"
8	p12 2.5H	"	"	"	"
9	p9 2.5A'	R <sub>1</sub> /Bal31	"	P9	R, Hc
10	p9 2.5B'	"	"	"	"
11	p9 2.5C'	"	"	"	"
12	M13mp8-HRO.9	HR <sub>1</sub> 0.9	p5RK'1	M13mp8	R, H
13	p12 0.9A	H/Bal31	M13mp8 HRO.9	p12	H, Hc
14	p12 HBg0.5	HBg 0.5	p12 0.9A	"	H, B
15	p12 D1	DD 0.45	p5RK'1	"	M
16	p12 D2	DD 0.35	"	"	"
17	p8 HPs1.5	HPs 1.5	P8H-H	p8	H, Ps
18	p9 D3	DD 0.45	p8HPs1.6	p9	Hc
19	p8 HR1.9	HR 1.9	P8H-H	p8	H, R
20	p9 1.9A	Bg/Bal31	p8HR1.9	p9	B, Hc
21	p13 PsH2.3	PsH 2.3	p8H-H	p13	Ps
22	p18 PsNs0.7	PsNs0.7	p13PsH2.3	p18	Ps, H

Se indican de forma abreviada los sitios de restricción flanqueantes de cada inserto, así como su tamaño en kpb. Los vectores utilizados en cada caso se denominan abreviadamente : p5=pBR325, p8=pUC8, p9=pUC9, etc. En último lugar, los sitios de restricción utilizados para la inserción en cada vector.



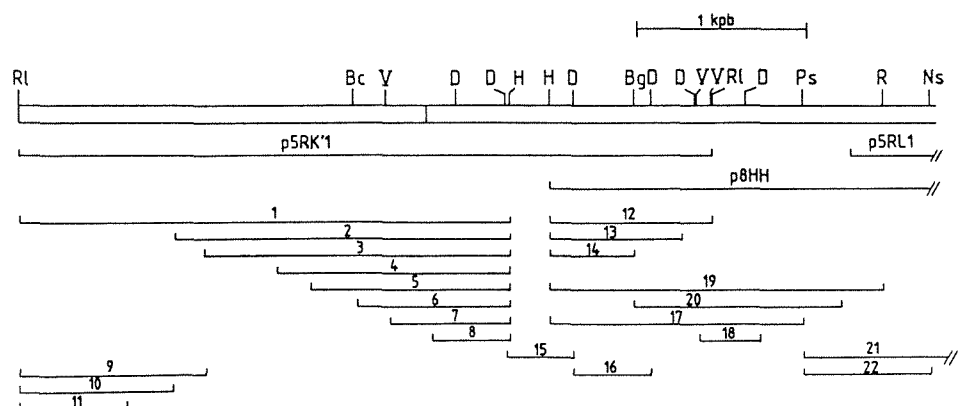


Figura 8. Subclones generados y utilizados para la secuenciación del fragmento terminal izquierdo.

Se representa la posición del inserto de cada subclon respecto a un mapa de restricción, en el que se señalan con trazos verticales los límites de la repetición terminal invertida. Cada subclon se identifica con el número correspondiente al asignado en la Tabla III. Se presentan asimismo las posiciones de los insertos de los recombinantes p5RK'1, p5RL2 y p8H-H.

TABLA IV. Subclones generados y utilizados para la secuenciación del fragmento terminal derecho.

N°	NOMBRE	INSERTO	PROCEDENCIA	VECTOR	SITIO INSERCIÓN
1	P9 MR3.4	MR <sub>1</sub> 3,4	p5RD'2	p9	Hc, R
2	p8 Bc1.6	Bc Bc1.6	p9MR3.4	p8	B
3	p9 1.6B	Ba131/Bc	p8Bc1.6	p9	Hc, R
4	p9 1.6C	"	"	"	"
5	p9 1.6D	"	"	"	"
6	p12MPvO.9	H PvO.9	p9MR3.4	p12	H, Hc
7	p12 O.9A	Ba131/H	p12MPvO.9	"	"
8	p12 O.9B	"	"	"	"
9	P9 V	MV-V R <sub>1</sub>	p9MR3.4	p9	Hc, R
10	p9 VA	Ba131/R <sub>1</sub>	p9 V	p9	Hc, R
11	p9 VB	"	"	"	"
12	p9 ScBcO.9	ScBcO.9	p8Bc1.6	"	Hc, B
13	p19 KM2.7	KM2.7	p5RD'2	p19	Hc, K
14	p19 SM1.5	SM1.5	"	"	Hc, S.
15	p12X 2.5	XX2.5	p5RD'2	p12	X
16	p9 N1.8	NN1.8	p12X2.5	p9	Hc
17	p12N	XN-NX	p12X2.5	p12	X
18	p5X	R <sub>1</sub> X-XR <sub>1</sub>	p5RD'2	p5	R

Las abreviaturas utilizadas son las mismas que en la tabla III.

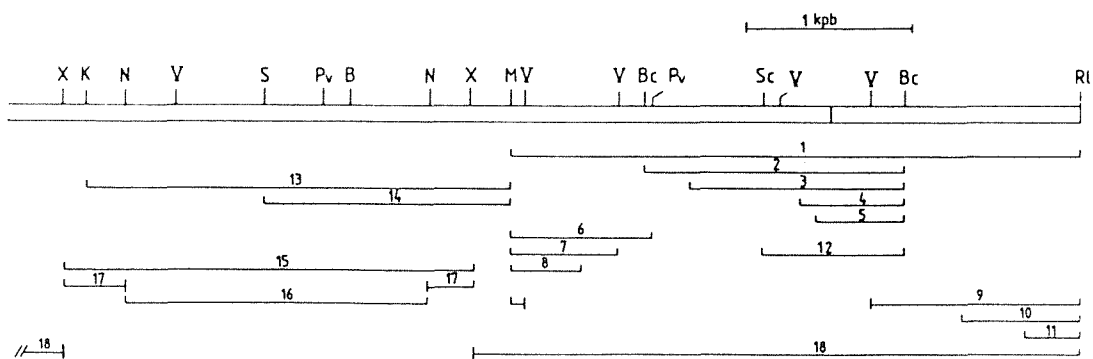


Figura 9. Subclones generados y utilizados para la secuenciación del fragmento terminal derecho.

Se representa la posición del inserto de cada subclon respecto a un mapa de restricción. Cada subclon se identifica con el número correspondiente al asignado en la tabla IV. La posición de la RTI se indica con trazos verticales en el mapa de restricción.

La detección de sitios DraI para secuenciación permitió la generación de algunos subclones (números 15, 16 y 18) muy útiles para solapar y orientar lecturas de secuencia previas.

La generación de subclones correspondientes al extremo más interno del fragmento EcoRI-K' y a la zona de EcoRI-L inmediatamente contigua se llevó a cabo partiendo del clon recombinante p8H-H, que contiene el inserto HindIII-HindIII de 3.9 kpb que solapa ambos fragmentos EcoRI, con lo que se disponía de las secuencias del fragmento EcoRI-K' orientadas hacia el interior del genoma que resultaron delecionadas durante la obtención de p5RK'1 (I. 2.4.). Se utilizaron sitios de restricción para generar subclones de esta zona salvo en el caso del recombinante p9 1.9A.

## 2.2.Fragmento terminal derecho (EcoRI-D').

La tabla IV muestra el procedimiento de obtención de los subclones utilizados en la secuenciación del extremo derecho. En la figura 9 se muestra la localización de los insertos de dichos subclones con respecto a un mapa de restricción de la zona terminal del fragmento EcoRI-D' clonado.

Tras el subclonaje del fragmento terminal SmaI-EcoRI<sub>1</sub> se detectó por mapeo de restricción la presencia en el mismo de dos sitios BclI. Dichos sitios se utilizaron para generar un subclon que incluía la parte inferior del fragmento (p8Bc 1.6), a partir del cual se generaron subclones utilizando el enzima Bal 31 (p9 1.6B, p9 1.6C y p9 1.6D).

El sitio PvuII detectado por secuenciación en el extremo izquierdo del fragmento BclI-BclI se utilizó para obtener un subclon que contuviera la parte izquierda del fragmento SmaI-EcoRI<sub>1</sub>, pl2MPv 0.9. Este subclon se utilizó a su vez para obtener subclones delecionados por medio de la nucleasa Bal 31: pl2 0.9A y pl2 0.9B.

La presencia de sitios EcoRV, detectados por secuenciación, en el fragmento SmaI-EcoRI<sub>1</sub> se aprovechó para generar un subclon que contuviera las secuencias situadas a la derecha del sitio BclI terminal (p9V). Análogamente al caso anterior se obtuvieron a partir de él subclones Bal 31 (p9VA y p9VB).

La presencia de sitios de restricción, tanto de enzimas para los que se habían elaborado mapas genómicos (I.2.4.) como los estudiados en el apartado anterior (ver tabla II), situados a la izquierda del sitio SmaI, permitieron el subclonaje directo de fragmentos de restricción, algunos de los cuales se confirmaron posteriormente por secuenciación (NcoI y PvuII).

### 3. ESTRATEGIA DE SECUENCIACION.

#### 3.1. Horquilla terminal.

En la secuenciación del fragmento terminal izquierdo se detectó un sitio de restricción BssHII (G/CGCGC) a aproximadamente 70 pb del extremo izquierdo del inserto del clon re-

combinante p5RK'1. Puesto que la nucleasa S1 utilizada en el clonaje de los fragmentos terminales había digerido en muy poca extensión el extremo izquierdo del fragmento EcoRI-K', este sitio BssHII debía encontrarse muy próximo al extremo izquierdo del genoma viral.

Para determinar las características de los fragmentos BssHII terminales se llevó a cabo un experimento cuyos pasos se esquematizan en la figura 10a. El resultado se muestra en la figura 10b.

Se observa la presencia de dos fragmentos que poseen una movilidad relativa muy similar, aunque no idéntica, correspondiente a un tamaño aproximado de 100 pb, que se denominaron lento y rápido ( l y r en la figura). Los dos fragmentos aparecen en las digestiones con BssHII tanto de los fragmentos EcoRI-K y K' como de los EcoRI-D y D'.

La intensidad de las bandas observadas en la autorradiografía correspondientes a los fragmentos lento y rápido es muy similar. Si la eficiencia de incorporación de radioactividad ha sido similar en ambos fragmentos, la proporción de fragmentos lento y rápido en cada extremo del genoma es aproximadamente la misma.

El procedimiento de secuenciación de los fragmentos lento y rápido se basó en el supuesto de que dichos fragmentos constituían una horquilla terminal que unía covalentemente ambos extremos del genoma. Por lo tanto, el marcaje terminal utilizando el sitio BssHII produciría fragmentos lento y rápido marcados terminalmente en un único extremo,

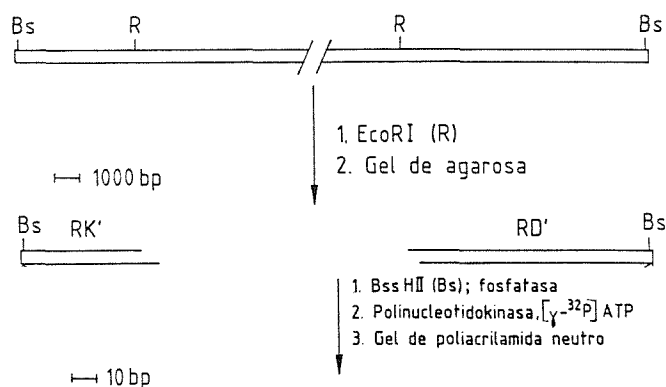
el interno, correspondiente al sitio de restricción, que podrían secuenciarse directamente tras su aislamiento y purificación.

El procedimiento seguido fue el esquematizado en la figura 11. El resultado se muestra en la figura 12. La presencia de una sola secuencia no palíndromica a partir de cada uno de los fragmentos indica que la suposición de partida de que los fragmentos lento y rápido forman una horquilla es correcta. Además, en ambos fragmentos, la secuencia más cercana al sitio BssHII (parte inferior del gel) es complementaria a la secuencia más alejada de dicho sitio (parte superior). Estas zonas de secuencia son idénticas en los fragmentos lento y rápido, mientras que la secuencia leída en la zona central del gel es distinta en ambos fragmentos (subrayada en la figura 12).

La secuencia próxima al sitio BssHII es idéntica a la existente entre el sitio EcoRI<sub>1</sub> terminal presente en el fragmento EcoRI-K' clonado y el sitio BssHII detectado en el mismo a una distancia de 64 pb de aquél y demuestra que los fragmentos lento y rápido se encuentran localizados en los extremos del genoma.

La conclusión es, por lo tanto, que los fragmentos lento y rápido constituyen las uniones covalentes terminales presentes en el genoma de VPPA, siendo su estructura la de una horquilla polinucleotídica.

a)



b)

pb

1

2

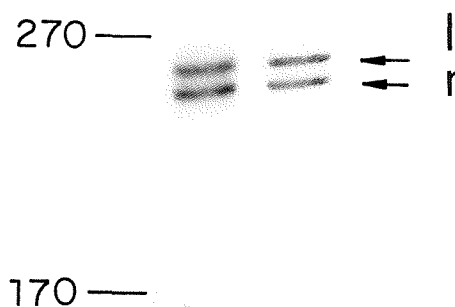


Figura 10. Fragmentos de restricción *Bss*HII lento y rápido. a) Se digirieron 4 ug de DNA de VPPA con *Eco*RI. Los fragmentos resultantes se sometieron a electroforesis en agarosa de bajo punto de fusión al 0.5 %. Los fragmentos *Eco*RI-K y K', y *Eco*RI-D y D' aislados del gel se desfosforilaron y marcaron radiactivamente con 5- $\gamma$ - $^{32}$ P-ATP y polinucleótidokinasa de T4, tras lo cual se les sometió a electroforesis en gel de poliacrilamida al 12 %. b) Se muestra la autorradiografía del gel seco, correspondiendo el canal 1 a los fragmentos D y D' y el 2 a los K y K'. l, fragmento lento; r, fragmento rápido



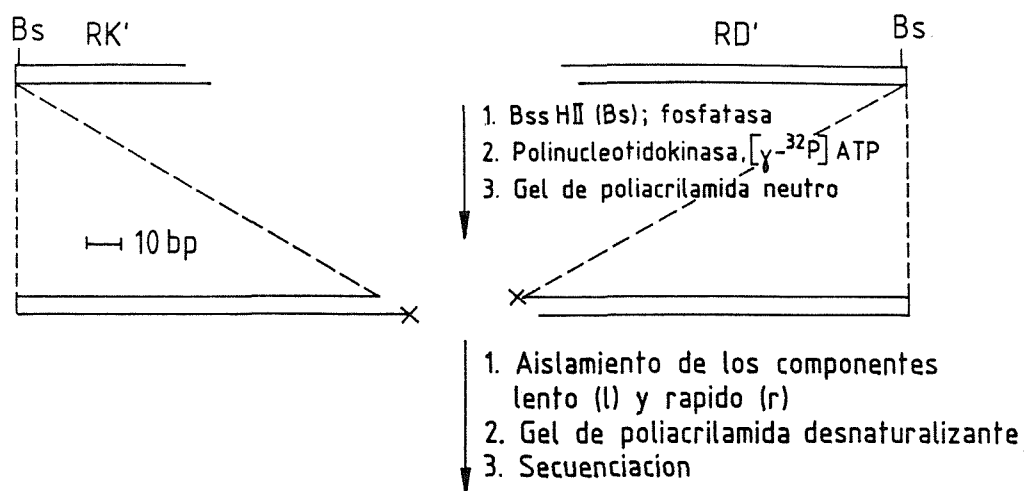


Figura 11. Aislamiento de los fragmentos BssHII lento y rápido marcados terminalmente para su secuenciación.

Se digirieron 8 ug de DNA de VPPA. Los fragmentos resultantes se desfosforilaron y marcaron terminalmente con polinucleótidokinasas de T4 y 5- $\gamma$ - $^{32}\text{P}$ -ATP (actividad específica, 5300 Ci mmol $^{-1}$ ) y se resolvieron por electroforesis en gel de poliacrilamida neutro (12 %). Los fragmentos lento y rápido se identificaron por autorradiografía y se sometieron por separado a electroforesis en un gel de poliacrilamida desnaturante (8 %). Los fragmentos se identificaron como en el caso anterior, se extrajeron por difusión y purificaron por cromatografía en minicolumna centrifugada y precipitación con etanol.

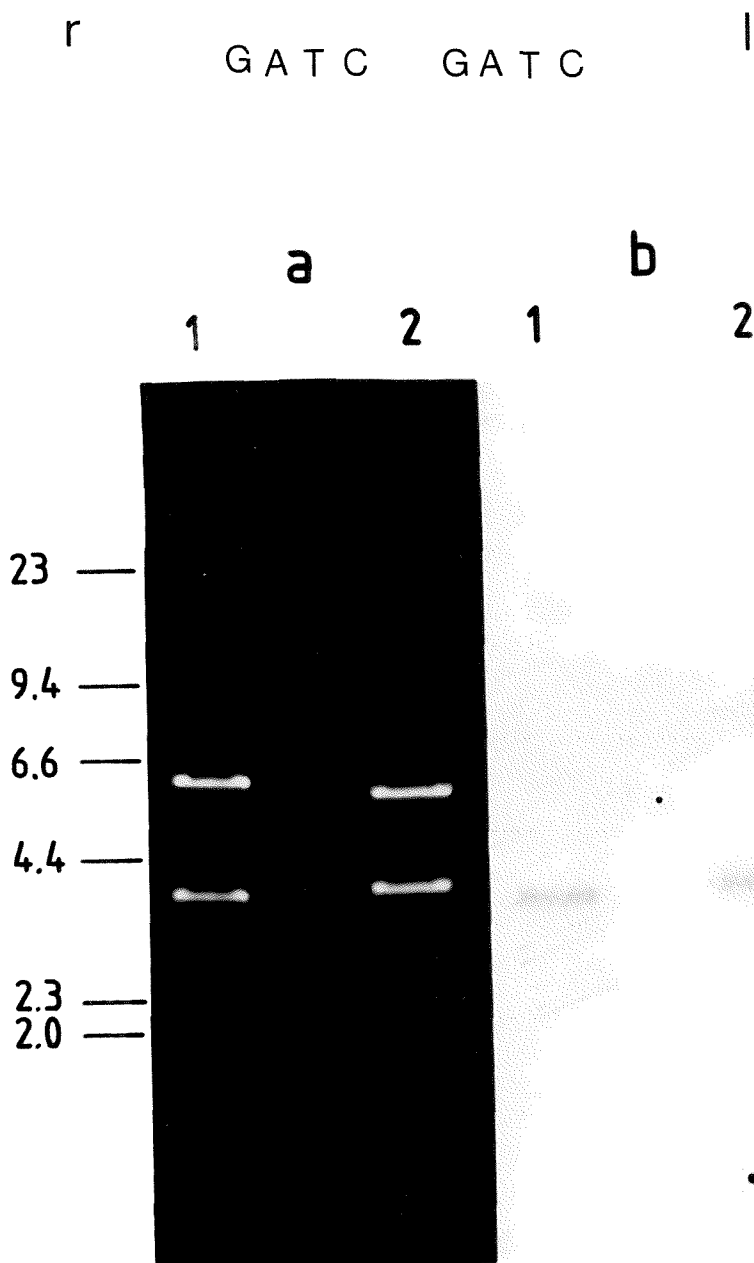


Figura 12. Secuencia de nucleótidos de los fragmentos de restricción lento y rápido.

Autorradiografía del gel de secuenciación (8 %) de dichos fragmentos. Los canales correspondientes a la distintas reacciones de degradación química se denotan por G(G), A(G+A), T(T+C) y C(C). Se muestra asimismo la secuencia de cada fragmento, correspondiendo el subrayado a las zonas en que ambas secuencias divergen.

### 3.2. Fragmento terminal izquierdo.

La estrategia de secuenciación empleada en la determinación de la estructura del fragmento terminal izquierdo se basó en los subclones generados según lo descrito en el apartado IV.2.1.

Los extremos a marcar radioactivamente fueron producidos por digestiones con enzimas de restricción que reconocían sitios presentes en los insertos de los clones o subclones correspondientes, o adyacentes a los insertos debido al proceso de clonaje en el sitio de inserción múltiple presente en los vectores de la serie pUC. El marcaje terminal de dichos extremos se realizó siguiendo los métodos ya descritos (II.16). El método seguido para la obtención de fragmentos marcados en un único extremo fue el que utiliza un segundo enzima de restricción, salvo en el caso del fragmento HindIII-HindIII de 239 pb, cuyas cadenas se separaron por electroforesis. Los fragmentos marcados se sometieron a las reacciones de degradación química y los productos de las mismas a electroforeesis en geles de secuenciación con el fin de obtener la secuencia de nucleótidos a partir del extremo marcado radioactivamente. La longitud de secuencia leída fue de 200-400 nucleótidos, dependiendo del experimento.

La figura 13 resume la estrategia de secuenciación empleada en el caso del fragmento terminal izquierdo, incluyendo el mapa de restricción de los enzimas relevantes para

extremo mercuriado común y extremos marcados distintos, cuya secuenciación origina lecturas solapantes.

Los otros dos tipos son aplicables a ambas reacciones de secuenciación, según el vector en el que se realice el subclonaje que implican. En los dos tipos se obtiene una colección de subclones con un extremo común y otro que presenta una delección de tamaño variable respecto al fragmento clonado original. La secuenciación utilizando estos subclones origina lecturas solapantes, con o sin un paso previo de identificación de subclones de tamaño progresivamente menor.

En uno de ellos, el primer paso es un corte al azar (con DNAsa I) y el segundo un corte con un enzima de restricción. Tras otros pasos adicionales para asegurar el ligamiento eficiente de los extremos, solo los recombinantes cuyo corte al azar se produjo en el inserto serán viables (109-111).

En el otro, tras una linearización del clon original con un enzima de restricción, las delecciones en el inserto se obtienen por una digestión progresiva con nucleasa ( Bal 31, Exo III + S1 o Exo III + Exo VII) y el extremo fijado se origina tras un segundo corte con enzima de restricción (112). Para asegurar la presencia de subclones con delecciones de la extensión adecuada se lleva a cabo una selección por tamaño, mediante electroforesis de los insertos deleccionados o de fagos recombinantes completos .

El aislamiento de DNA molde monocatenario necesario para las reacciones enzimáticas de secuenciación ha sido faci-

litado por la construcción de vectores de clonaje derivados del bacteriófago M13 (113). Los fragmentos de DNA bicatenario de interés se pueden introducir en la forma replicativa bicatenaria de los fagos vectores. Tras la transfección de las células huésped adecuadas, el fragmento de DNA exógeno ligado a la cadena (+) de la forma replicativa es empaquetado en forma de fago recombinante cuyo genoma es DNA circular monocatenario.

Messing y colaboradores han desarrollado una serie de vectores (M13mp) que contienen un sitio de clonaje múltiple y un marcador genético eficiente que permite la discriminación entre fagos recombinantes y no recombinantes (113,114). La región promotora/operadora del operón de la lactosa y la secuencia que codifica los 145 aminoácidos N-terminales de la  $\beta$ -galactosidasa se han introducido en una región no esencial del genoma de M13. Si dichos fagos infectan un huésped E. coli con un gen defectivo de la  $\beta$ -galactosidasa (que carece de los aminoácidos 11-41), la  $\beta$ -complementación da lugar a actividad  $\beta$ -galactosidasa, que se detecta fácilmente en presencia del substrato X-gal, cuyo producto de hidrólisis es azul. La síntesis de  $\beta$ -galactosidasa puede ser incrementada por la acción del inductor IPTG. Las bacterias infectadas con fagos recombinantes que contienen un fragmento de DNA exógeno en el sitio de clonaje múltiple localizado en el gen de la  $\beta$ -galactosidasa son incapaces de complementar y son, por lo tanto fácilmente detectables.

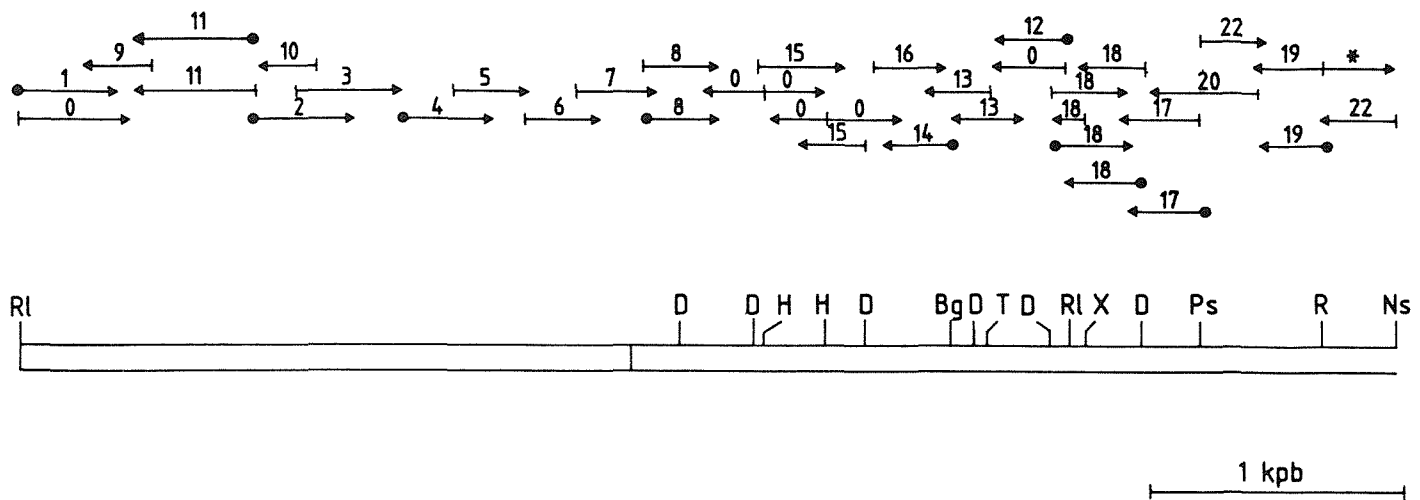


Figura 13. Estrategia de secuenciación del fragmento terminal izquierdo.

Se representan las lecturas a partir de los extremos marcados por medio de flechas. Los símbolos  $\leftarrow$  y  $\bullet$  en el origen de cada flecha indican marcajes 5' y 3' terminales, respectivamente. El número sobre cada flecha denota el subclon utilizado (tabla III) para la obtención de los fragmentos marcados terminalmente. El número 0 corresponde al clon original, p5RK'l, y el símbolo \* a p5RLl. Como referencia se presenta asimismo el mapa de restricción de la zona terminal, señalándose la posición de la RTI con trazos verticales.

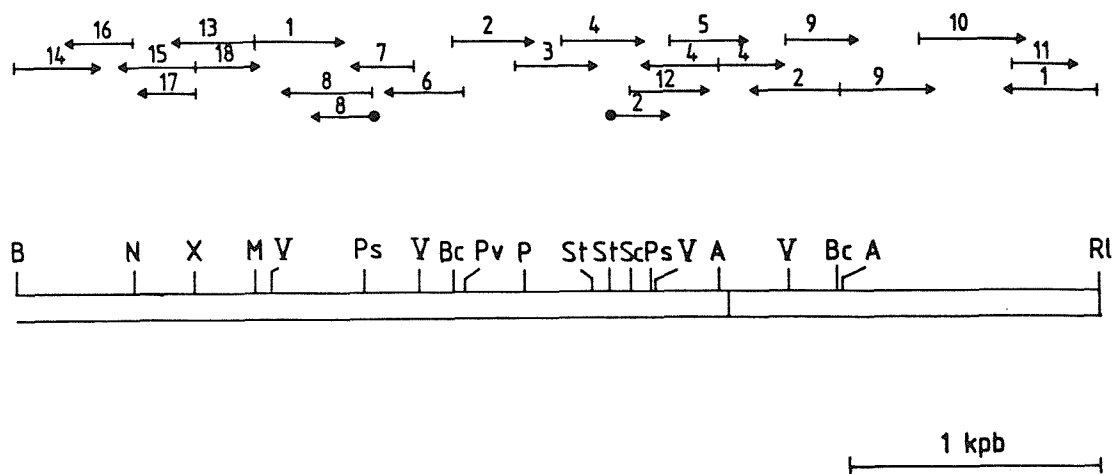


Figura 14. Estrategia de secuenciación del fragmento terminal derecho.

Los símbolos son los mismos que los utilizados en la figura 13.

La secuenciación de la parte de la repetición terminal derecha presente en el recombinante p5RD'2 se apoyó tanto en sitios de restricción presentes en la repetición (EcoRV, BclI), como en subclones delecionados obtenidos utilizando la nucleasa Bal 31.

#### 4. SECUENCIA DE NUCLEOTIDOS.

##### 4.1. Horquilla terminal.

La unión covalente entre las dos cadenas del DNA del VPPA presente en ambos extremos del genoma consiste en dos secuencias, denominadas lenta y rápida, de acuerdo con la movilidad electroforética de los fragmentos de restricción BssHII terminales a partir de los cuales se determinó su secuencia.

Según se muestra en la figura 12, la secuencia leída a partir del sitio BssHII es distinta en los fragmentos lento y rápido entre los nucleótidos 71 y 107, numerados tomando como referencia el extremo 5' del sitio de restricción BssHII (secuencia subrayada). A partir del nucleótido 108, la secuencia vuelve a ser idéntica en ambos fragmentos y perfectamente complementaria a la de los 70 primeros nucleótidos.

Si se comparan las zonas divergentes (de 37 nucleótidos de longitud) de las secuencias lenta y rápida con polaridades opuestas son invertidas y complementarias, como se observa en la figura 15.



Otro rasgo significativo de dichas zonas es su imposibilidad para adoptar una estructura secundaria con total apareamiento de bases. La estructura secundaria más estable desde un punto de vista termodinámico se muestra asimismo en la figura 15.

Las estructuras obtenidas se muestran en la figura 15. Contienen 24 bases apareadas, siendo todos los apareamientos A-T y 13 bases no apareadas, 5 de las cuales se encuentran en la parte más terminal. Esta estructura tiene un alto contenido en A+T: 36 de 37 nucleótidos, lo que equivale a un 97%.

#### 4.2. Repetición terminal invertida.

La secuenciación del fragmento EcoRI-K' y de la zona terminal del fragmento EcoRI-D' clonado, confirmó la presencia en los extremos del genoma de una repetición terminal invertida detectada previamente por experimentos de hibridación y microscopía electrónica (143).

La repetición terminal invertida presente en el extremo izquierdo se secuenció completamente ya que la secuencia más terminal obtenida a partir del fragmento EcoRI-K' clonado se solapó con la secuencia terminal obtenida a partir del genoma completo utilizando el sitio BssHII (IV.3.1). Debido a la delección terminal generada durante el clonaje del fragmento EcoRI-D', la repetición del extremo derecho se secuenció solo parcialmente.

La secuencia de la repetición presente en el extremo izquierdo se muestra en la figura 16. Consta de 2440 nucleó-

	1	10	20	30	37	
	5'-ATATATATATAAAATTATAAAATATATAATATACTTATA-3'					
r	5'-TATAAGTATATTATATATTTTATAAATTTTATATATAT-3'					
	3'-ATATTCATATAATATATAAAATATTAAATATATATA-5'					

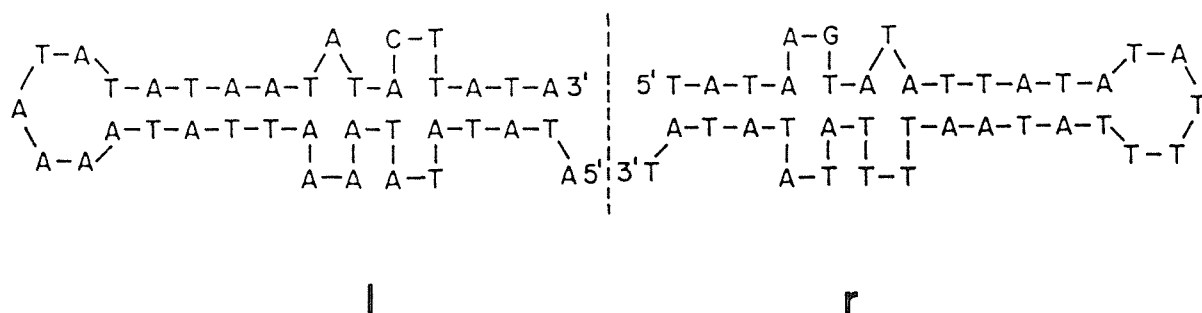


Figura 15. Secuencia de nucleótidos y estructura secundaria de la horquilla terminal.

En la parte superior se muestran las secuencias de la horquilla terminal presentes en los fragmentos BssHII lento y rápido (l y r). Se representa asimismo la secuencia l con polaridad opuesta. En la parte inferior se muestran las estructuras secundarias más estables termodinámicamente.

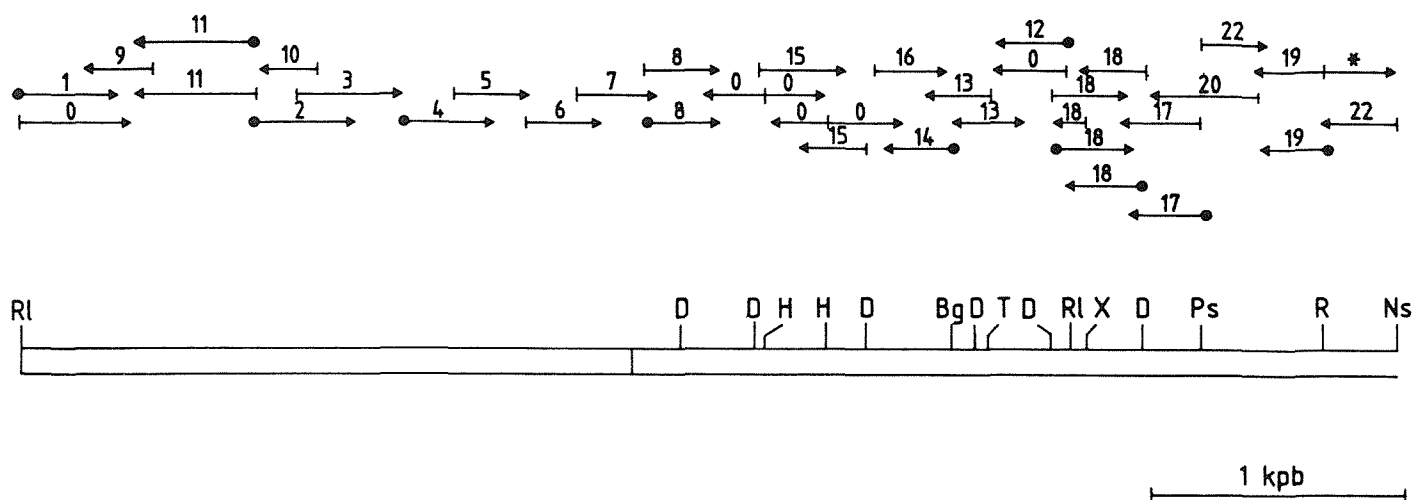


Figura 13. Estrategia de secuenciación del fragmento terminal izquierdo.

Se representan las lecturas a partir de los extremos marcados por medio de flechas. Los símbolos  $\vdash$  y  $\bullet$  en el origen de cada flecha indican marcajes 5' y 3' terminales, respectivamente. El número sobre cada flecha denota el subclon utilizado (tabla III) para la obtención de los fragmentos marcados terminalmente. El número 0 corresponde al clon original, p5RK'l, y el símbolo \* a p5RLl. Como referencia se presenta asimismo el mapa de restricción de la zona terminal, señalándose la posición de la RTI con trazos verticales.

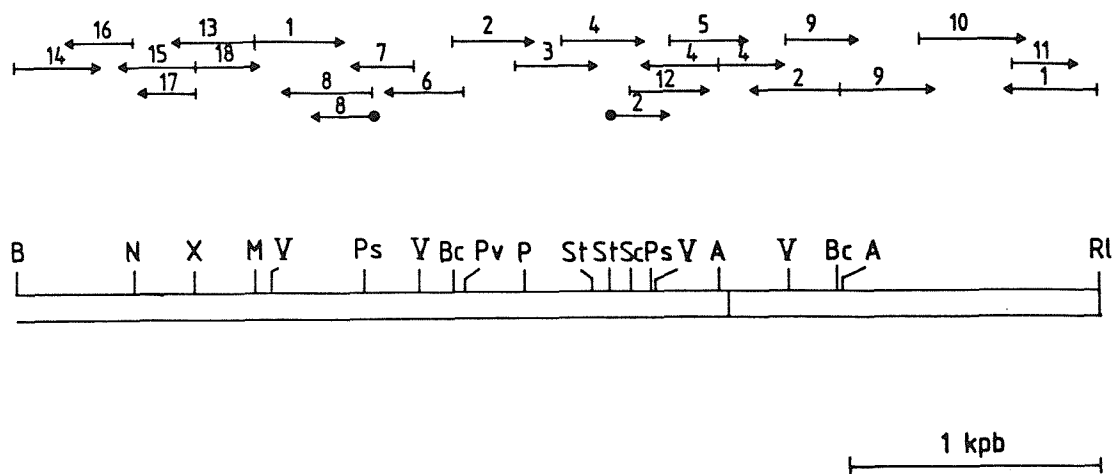


Figura 14. Estrategia de secuenciación del fragmento terminal derecho.

Los símbolos son los mismos que los utilizados en la figura 13.

La secuenciación de la parte de la repetición terminal derecha presente en el recombinante p5RD'2 se apoyó tanto en sitios de restricción presentes en la repetición (EcoRV, BclI), como en subclones delecionados obtenidos utilizando la nucleasa Bal 31.

#### 4. SECUENCIA DE NUCLEOTIDOS.

##### 4.1. Horquilla terminal.

La unión covalente entre las dos cadenas del DNA del VPPA presente en ambos extremos del genoma consiste en dos secuencias, denominadas lenta y rápida, de acuerdo con la movilidad electroforética de los fragmentos de restricción BssHII terminales a partir de los cuales se determinó su secuencia.

Según se muestra en la figura 12, la secuencia leída a partir del sitio BssHII es distinta en los fragmentos lento y rápido entre los nucleótidos 71 y 107, numerados tomando como referencia el extremo 5' del sitio de restricción BssHII (secuencia subrayada). A partir del nucleótido 108, la secuencia vuelve a ser idéntica en ambos fragmentos y perfectamente complementaria a la de los 70 primeros nucleótidos.

Si se comparan las zonas divergentes (de 37 nucleótidos de longitud) de las secuencias lenta y rápida con polaridades opuestas son invertidas y complementarias, como se observa en la figura 15.

Otro rasgo significativo de dichas zonas es su imposibilidad para adoptar una estructura secundaria con total apareamiento de bases. La estructura secundaria más estable desde un punto de vista termodinámico se muestra asimismo en la figura 15.

Las estructuras obtenidas se muestran en la figura 15. Contienen 24 bases apareadas, siendo todos los apareamientos A-T y 13 bases no apareadas, 5 de las cuales se encuentran en la parte más terminal. Esta estructura tiene un alto contenido en A+T: 36 de 37 nucleótidos, lo que equivale a un 97%.

#### 4.2. Repetición terminal invertida.

La secuenciación del fragmento EcoRI-K' y de la zona terminal del fragmento EcoRI-D' clonado, confirmó la presencia en los extremos del genoma de una repetición terminal invertida detectada previamente por experimentos de hibridación y microscopía electrónica (143).

La repetición terminal invertida presente en el extremo izquierdo se secuenció completamente ya que la secuencia más terminal obtenida a partir del fragmento EcoRI-K' clonado se solapó con la secuencia terminal obtenida a partir del genoma completo utilizando el sitio BssHII (IV.3.1). Debido a la delección terminal generada durante el clonaje del fragmento EcoRI-D', la repetición del extremo derecho se secuenció solo parcialmente.

La secuencia de la repetición presente en el extremo izquierdo se muestra en la figura 16. Consta de 2440 nucleó-

	1	10	20	30	37	
	5'-ATATATATATAAAATTATAAAATATATAATATACTTATA-3'					
r	5'-TATAAGTATATTATATATTTTATAAATTTTATATATAT-3'					
	3'-ATATTCATATAATATATAAAATATTAAATATATATA-5'					

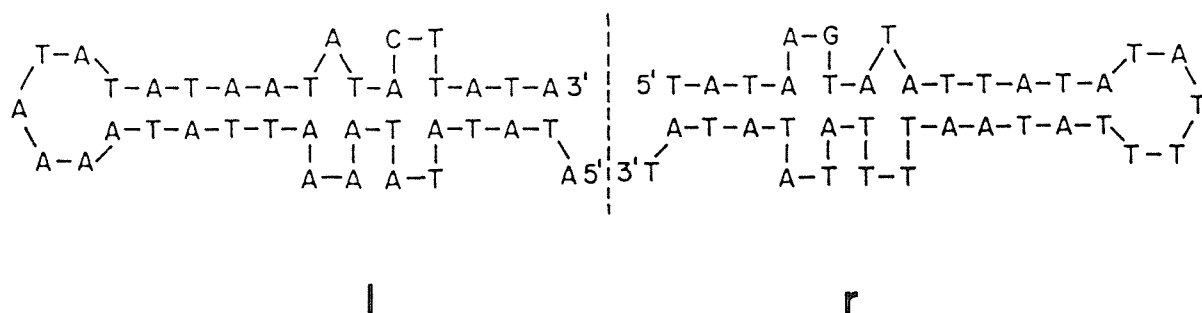


Figura 15. Secuencia de nucleótidos y estructura secundaria de la horquilla terminal.

En la parte superior se muestran las secuencias de la horquilla terminal presentes en los fragmentos BssHII lento y rápido (l y r). Se representa asimismo la secuencia l con polaridad opuesta. En la parte inferior se muestran las estructuras secundarias más estables termodinámicamente.

	AAAGCAGGTA	CAATTTTATT	ATATAGTGCA	GATCTTTTTT	GTGAAAAAA	AATAATTATT	60
RD27	TTTTTGCGCG	CCGGAATAAT	TATTTTTTT	GCGCGGAAAT	ATTTTTGCAG	TGGCGATTGA	120
RD35	TTTTTTTTTC	AGTTGTCTGC	CCGTGGCGGA	ATAATTATTT	TTTCGCCGGA	AAGCATTTTG	180
RD35, RD27	TAGTTTCACA	AGATGCAGAA	TTTGGCGGGC	CGGAATAATT	ATTTTTTCGC	CGGAAAGTAT	240
	TTTGTGTACC	G TAGTTTCAC	TAGATTTTCGC	GGTGCATACC	GGTGGGCGAT	AGGTATTAAA	300
RDT34	ATAATTTGTA	CTGCTGTAGG	CGTTAAAGAT	TAAAATTATT	ATTACTGCTG	TAGGCGTTAA	360
	ACATTAAACT	TATTACTACT	GCTATAGGCG	TTAAACATTA	AAATTATTAC	TACTGTTGTA	420
	GGCGTTAAAC	ATTAAAATTA	TTACTACTGC	TGTAGGCGTC	AAAGATTAAA	ATTATTACTA	480
	CTGCTGTAGG	CGTTAAACAT	TAAACTTATT	ACTACTGCTA	TAGGCGTTAA	ACATTAAAAT	540
	TATTACTACT	GTTGTAGGCG	TCAAAGATTA	AAATTATTAC	TACTGCTGTA	GGCGTTAAAC	600
	ATTAAAATTA	TTACTACTGC	TGTAGGCGTT	AAACATTAAA	ATTATTACTA	CTGTTGTAGG	660
	CGTTAAACAT	TAAAATTATT	ACTACTGTTG	TAGGCGTCAA	AGATTAAAAT	TATTACTACT	720
	GCTGTAGGCG	TTAAACATTA	AAATTATTAC	TACTGCTGTA	GGCGTTAAAC	ATTAAAATTA	780
	TTACTACTGT	TGTAGGCGTT	AAACATTAAA	ATTATTACTA	CTGCTGTAGG	CGTCAAAGAT	840
	TAAAATTATT	ACTACTGCTG	TAGGCGTTAA	ACATTAAACT	TATTACTACT	GCTATAGGCG	900
	TTAAACATTA	AACTTATTAC	TACTGCTATA	GGCGTTAAAC	ATTAAAATTA	TTACTACTGC	960
	TGTAGGCGTT	AAACATTAAA	ATTATTACTA	CTGTTGTAGG	CGTCAAACAT	TAAAATTATT	1020
	ACTACTGTTG	TAGGCGTTAA	ACATTAAAAT	TATTACTACT	GCTGTAGGCG	TTAAACATTA	1080
	AACTTATTAC	TACTGCTATA	GGCGTTAAAC	ATTAAAATTA	TTACTACTGT	TGTAGGCGTT	1140
	AAACATTAAA	ATTATTACTA	CTGCTGTAGG	CGTCAAAGAT	TAAAATTATT	ACTACTGCTG	1200
	TAGGCGTTAA	ACATTAAACT	TATTACTACT	GCTATAGGCG	TTAAACATTA	AAATTATTAC	1260
	TACTGTTGTA	GGCGTCAAAG	ATTAAAATTA	TTACTACTGC	TGTAGGCGTT	AAACATTAAA	1320
	ATTATTACTA	CTGCTGTAGG	CGTTAAACAT	TAAAATTATT	ACTACTGTTG	TAGGCGTTAA	1380
	ACATTAAAAT	TATTACTACT	GTTGTAGGCG	TCAAAGATTA	AAATTATTAC	TACTGCTGTA	1440
	GGCGTTAAAC	ATTAAAATTA	TTACTACTGC	TGTAGGCGTT	AAACATTAAA	ATTATTACTA	1500
	CTGTTGTAGG	CGTTAAACAT	TAAAATTATT	ACTACTGCTG	TAGGCGTCAA	AGATTAAAAT	1560
	TATTACTACT	GCTGTAGGCG	TTAAACATTA	AACTTATTAC	TACTGCTATA	GGCGTTAAAC	1620
	ATTAAAATTA	TTACTACTGT	TGTAGGCGTC	AAAGATTAAA	ATTATTACTA	CTGCTGTAGG	1680
	CGTTAAACAT	TAAAATTATT	ACTACTGCTG	TAGGCGTTAA	ACATTATTAC	CGCTGTAGGC	1740
	G TAGCTAAAT	TTTAATGTTT	TTTCGTAAAA	AATCGCCAGT	CACAAAAAA	AGCAGCCATG	1800
	TATATCTGTG	ATGTTTGGTG	TTTTATCTAC	ATGCTATTGG	CATTCCGTGAT	ATTCATGGTA	1860
	AAGTGCTATA	CTATAACCAT	CTTATGTGGG	CTCTTTATAC	ATCTTATTGG	CAGGAGGCC	1920
	ACTATATATA	TCCTGTTAAG	TAGTTTAGCT	GTATACTATT	TAATGTTTTT	CGCTTGATCA	1980
RDT27	AGAAATATAC	AAAAATCTAA	GTGCACCATT	TTATGGGCAT	CAATCAATCA	CAGCATTCTA	2040
	TGGGCGTGAA	TCAATCACAC	CATTTTATGG	GCGTCAATCA	ATCACACCAT	TTTATGGGCA	2100
	TCAATCAATC	ACAGCATTCT	ATGGGCGTGA	ATCAACACAT	TATTGGGGAC	AGCTACCTGA	2160
RDT33	TATCGTGTAT	GGTGGTGGCA	CCGTGTTGTT	TGAAGGCGTG	TATGGTATTA	GCGCCATGTT	2220
	GTTTGAAGGC	GTGTATGGTA	TTAGCGCCAT	GTTGTTTGAA	GGCAGCGCGT	TAGTAACAAC	2280
	GTTTTTTTTC	GTATGACAAC	AGATATGTAT	AAGATAGCAA	ATTAATGCTA	TTGTCATCAA	2340
	TGAGAAACAT	AAGTAGATAT	AGACCTGCCT	CTTGGTAGTC	CAATGATCCA	TGTTGCACCT	2400
	AAAAATCCCA	AAAAAAAACA	TTAAGTTGTG	GAGGGTAAGA			

Figura 16. Secuencia de la repetición terminal invertida presente en el extremo izquierdo.

La orientación de la secuencia es la convencional (5'-3'), correspondiendo los primeros nucleótidos a los más próximos a los extremos del genoma. La posición de los distintos tipos de repeticiones se denota por símbolos:  $\Delta$ , RD27;  $\diamond$ , RD35;  $\blacktriangle$ , RDT34;  $\star$ , RDT27;  $\blacklozenge$ , RDT33.



	AAAGCAGGTA	CAATTTTATT	ATATAGTGCA	GATCTTTTTT	GTGAAAAAAA	AATAATTATT	60
<b>RD27</b>	TTTTTGCGCG	CCGGAATAAT	TATTTTTTTT	GCGCGGAAAT	ATTTTTGCAG	TGGCGATTGA	120
<b>RD35</b>	TTTTTTTTTC	AGTTGTCTGC	CCGTGGCGGA	ATAATTATTT	TTTCGCCCGGA	AAGCATTTTG	180
<b>RD35, RD27</b>	TAGTTTCACA	AGATGCAGAA	TTTGGCGGGC	CGGAATAATT	ATTTTTTCGC	CGGAAAGTAT	240
	TTTGTGTACC	GTAGTTTCAC	TAGATTTTCGC	GGTGCATACC	GGTGGGCGAT	AGGTATTAAA	300
<b>RDT34</b>	ATAATTTGTA	CTGCTGTAGG	CGTTAAAGAT	TAAAATTATT	ATTACTGCTG	TAGGCGTTAA	360
	ACATTAAACT	TATTACTACT	GCTATAGGCG	TTAAACATTA	AAATTATTAC	TACTGTTGTA	420
	GGCGTTAAAC	ATTAAAATTA	TTACTACTGC	TGTAGGCGTC	AAAGATTAAA	ATTATTACTA	480
	CTGCTGTAGG	CGTTAAACAT	TAAACTTATT	ACTACTGCTA	TAGGCGTTAA	ACATTAAAAT	540
	TATTACTACT	GTTGTAGGCG	TCAAAGATTA	AAATTATTAC	TACTGCTGTA	GGCGTTAAAC	600
	ATTAAAATTA	TTACTACTGC	TGTAGGCGTT	AAACATTAAA	ATTATTACTA	CTGTTGTAGG	660
	CGTTAAACAT	TAAAATTATT	ACTACTGTTG	TAGGCGTCAA	AGATTAAAAT	TATTACTACT	720
	GCTGTAGGCG	TTAAACATTA	AAATTATTAC	TACTGCTGTA	GGCGTTAAAC	ATTAAAATTA	780
	TTACTACTGT	TGTAGGCGTT	AAACATTAAA	ATTATTACTA	CTGCTGTAGG	CGTCAAAGAT	840
	TAAAATTATT	ACTACTGCTG	TAGGCGTTAA	ACATTAAACT	TATTACTACT	GCTATAGGCG	900
	TTAAACATTA	AACTTATTAC	TACTGCTATA	GGCGTTAAAC	ATTAAAATTA	TTACTACTGC	960
	TGTAGGCGTT	AAACATTAAA	ATTATTACTA	CTGTTGTAGG	CGTCAAACAT	TAAAATTATT	1020
	ACTACTGTTG	TAGGCGTTAA	ACATTAAAAT	TATTACTACT	GCTGTAGGCG	TTAAACATTA	1080
	AACTTATTAC	TACTGCTATA	GGCGTTAAAC	ATTAAAATTA	TTACTACTGT	TGTAGGCGTT	1140
	AAACATTAAA	ATTATTACTA	CTGCTGTAGG	CGTCAAAGAT	TAAAATTATT	ACTACTGCTG	1200
	TAGGCGTTAA	ACATTAAACT	TATTACTACT	GCTATAGGCG	TTAAACATTA	AAATTATTAC	1260
	TACTGTTGTA	GGCGTCAAAG	ATTAAAATTA	TTACTACTGC	TGTAGGCGTT	AAACATTAAA	1320
	ATTATTACTA	CTGCTGTAGG	CGTTAAACAT	TAAAATTATT	ACTACTGTTG	TAGGCGTTAA	1380
	ACATTAAAAT	TATTACTACT	GTTGTAGGCG	TCAAAGATTA	AAATTATTAC	TACTGCTGTA	1440
	GGCGTTAAAC	ATTAAAATTA	TTACTACTGC	TGTAGGCGTT	AAACATTAAA	ATTATTACTA	1500
	CTGTTGTAGG	CGTTAAACAT	TAAAATTATT	ACTACTGCTG	TAGGCGTCAA	AGATTAAAAT	1560
	TATTACTACT	GCTGTAGGCG	TTAAACATTA	AACTTATTAC	TACTGCTATA	GGCGTTAAAC	1620
	ATTAAAATTA	TTACTACTGT	TGTAGGCGTC	AAAGATTAAA	ATTATTACTA	CTGCTGTAGG	1680
	CGTTAAACAT	TAAAATTATT	ACTACTGCTG	TAGGCGTTAA	ACATTATTAC	CGCTGTAGGC	1740
	GTAGCTAAAT	TTTAATGTTT	TTTCGTAAAA	AATCGCCAGT	CACAAAAAAA	AGCAGCCATG	1800
	TATATCTGTG	ATGTTTGGTG	TTTTATCTAC	ATGCTATTGG	CATTCCTGAT	ATTCATGGTA	1860
	AAGTGCTATA	CTATAACCAT	CTTATGTGGG	CTCTTTATAC	ATCTTATTGG	CAGGAGGCCC	1920
	ACTATATATA	TCCTGTTAAG	TAGTTTAGCT	GTATACTATT	TAATGTTTTT	CGCTTGATCA	1980
<b>RDT27</b>	AGAAATATAC	AAAAATCTAA	GTGCACCAT	TTATGGGCAT	CAATCAATCA	CAGCATTCTA	2040
	TGGGCGTGAA	TCAATCACAC	CATTTTATGG	GCGTCAATCA	ATCACACCAT	TTTATGGGCA	2100
	TCAATCAATC	ACAGCATTCT	ATGGGCGTGA	ATCAACACAT	TATTGGGGAC	AGCTACCTGA	2160
<b>RDT33</b>	TATCGTGTAT	GGTGGTGGCA	CCGTGTTGTT	TGAAGGCGTG	TATGGTATTA	GCGCCATGTT	2220
	GTTTGAAGGC	GTGTATGGTA	TTAGCGCCAT	GTTGTTTGAA	GGCAGCGCGT	TAGTAACAAC	2280
	GTTTTTTTTTC	GTATGACAAC	AGATATGTAT	AAGATAGCAA	ATTAATGCTA	TTGTCATCAA	2340
	TGAGAAACAT	AAGTAGATAT	AGACCTGCCT	CTTGGTAGTC	CAATGATCCA	TGTTGCACCT	2400
	AAAAATCCCA	AAAAAAAACA	TTAAGTTGTG	GAGGGTAAGA			

Figura 16. Secuencia de la repetición terminal invertida presente en el extremo izquierdo.

La orientación de la secuencia es la convencional (5'-3'), correspondiendo los primeros nucleótidos a los más próximos a los extremos del genoma. La posición de los distintos tipos de repeticiones se denota por símbolos:  $\Delta$ , RD27;  $\diamond$ , RD35;  $\blacktriangle$ , RDT34;  $\star$ , RDT27;  $\blacklozenge$ , RDT33.

tidos. La parte de la secuencia determinada en el extremo derecho se muestra en la figura 17. De sus 1525 nucleótidos, cuatro son diferentes de los presentes en el extremo izquierdo.

El análisis de la secuencia de la repetición terminal invertida muestra que está constituida a su vez por secuencias repetidas separadas por secuencias únicas (es decir, repetidas solo en el otro extremo). Las secuencias repetidas son de dos tipos: directas y directas en tandem. Todas ellas tienen un tamaño unidad de alrededor de 30 nucleótidos. En las figuras 16 y 17 se muestran señaladas con símbolos todos estos tipos de repeticiones.

El comienzo de la repetición terminal invertida consta de 294 nucleótidos en los que se alternan segmentos de secuencia única con secuencias repetidas directas (RD27 y RD35). A continuación se encuentran 1428 nucleótidos que constituyen el mayor bloque de repeticiones en tandem, 42 repeticiones de 34 nucleótidos (RDT34) a los que siguen 283 nucleótidos de secuencia única. Tras estos, dos pequeños grupos de repeticiones directas en tandem (RDT27 y RDT33), de  $4 \times 27 + 1 \times 26$  (133) y  $3 \times 33$  (99) nucleótidos separados por 25 nucleótidos de secuencia única. Finalmente, 177 nucleótidos de secuencia única. Esta disposición se esquematiza en la figura 18.

Todas estas repeticiones son imperfectas, es decir, presentan microheterogeneidad en su secuencia, pero restringida a un reducido número de posiciones. En la figura 18 se

muestran las secuencias consenso de las repeticiones, así como sus posiciones variables. En el caso de las RDT27, la quinta repetición contiene una base menos. Como se observa en dicha figura 18 los 17 últimos nucleótidos de la repetición directa de 27 nucleótidos (RD27) son idénticos a los 17 primeros nucleótidos de la repetición directa de 35 nucleótidos (RD35). Si se representa por B esta secuencia común y por A y C las secuencias no comunes de RD27 y RD35, respectivamente la estructura de los 294 primeros nucleótidos se puede representar como

-----AB-----BC-----ABC-----

donde --- denota las zonas de secuencia única.

La composición de todas estas repeticiones obtenida a partir de su secuencia consenso es, desde moderadamente rica en A+T (RDT 33, 54%), hasta muy rica en A+T (RDT34, 73%). Este alto contenido en A+T no es solo distintivo de la repetición terminal invertida (67.5%), sino del fragmento EcoRI-K' completo (68.5%).

Los nucleótidos diferentes en ambos extremos forman parte todos ellos de repeticiones en tandem (RDT34). En la tabla V se resumen las bases distintas y sus posiciones.

Las microheterogeneidades de secuencia entre las repeticiones son restringidas. En primer lugar, se producen en un reducido número de posiciones (2 y 1 en RD27 y RD35; 4, 5 y 8 en RDT27, RDT33 y RDT34, respectivamente). Además, en dichas posiciones solo existen dos bases alternativas en todos los casos, salvo en la posición 14 de RDT34. Este último he-

**RDT34** ATTACTACTG CTGTAGGCGT TAAACATTAA AATTATTACT ACTGTTGTAG GCGTTAAACA 60  
TTAAAATTAT TACTACTGTT GTAGGCGTCA AAGATTAAAA TTATTACTAC TGCTGTAGGC 120  
GTTAAACATT AAAATTATTA CTA CTACTGCTGT AGGCGTTAAA CATTAAACTT ATTACTACTG 180  
CTATAGGCGT TAAACATTAA AATTATTACT ACTGTTGTAG GCGTTAAACA TTAAAATTAT 240  
TACTACTGCT GTAGGCGTCA AAGATTAAAA TTATTACTAC TGCTGTAGGC GTTAAACATT 300  
AAACTTATTA CTACTGCTAT AGGCGTTAAA CATTAAAATT ATTACTACTG TTGTAGGCGT 360  
CAAAGATTAA AATTATTACT ACTGCTGTAG GCGTTAAACA TTAAAATTAT TACTACTGCT 420  
GTAGGCGTTA AACATTAAAA TTATTACTAC TGTTGTAGGC GTTAAACATT AAAATTATTA 480  
CTACTGTTGT AGGCGTCAAA GATTAAAATT ATTACTACTG CTGTAGGCGT TAAACATTAA 540  
AATTATTACT ACTGCTGTAG GCGTTAAACA TTAAAATTAT TACTACTGTT GTAGGCGTTA 600  
AACATTAAAA TTATTACTAC TGCTGTAGGC GTCAAAGATT AAAATTATTA CTACTGCTGT 660  
AGGCGTTAAA CATTAAACTT ATTACTACTG CTATAGGCGT TAAACATTAA AATTATTACT 720  
ACTGTTGTAG GCGTCAAAGA TTAAAATTAT TACTACTGCT GTAGGCGTTA AACATTAAAA 780  
TTATTACTAC TGCTGTAGGC GTTAAACATT ATTACCGCTG TAGGCGTAGC TAAATTTTAA 840  
TGTTTTTTTCG TAAAAAATCG CCAGTCACAA AAAAAAGCAG CCATGTATAT CTGTGATGTT 900  
TGGTGTTTTA TCTACATGCT ATTGGCATT C TGATATTCA TGGTAAAGTG CTATACTATA 960  
ACCATCTTAT GTGGGCTCTT TATACATCTT ATTGGCAGGA GGCCCACTAT ATATATCCTG 1020  
TTAAGTAGTT TAGCTGTATA CTATTTAATG TTTTTCGCTT GATCAAGAAA TATACAAAAA 1080  
**RDT27** TCTAAGTGCA CCATTTTATG GGCATCAATC AATCACAGCA TTCTATGGGC GTGAATCAAT 1140  
CACACCATTT ★ TATGGGCGTC AATCAATCAC ACCATTTTAT ★ GGCATCAAT CAATCACAGC 1200  
**RDT33** ★ ATTCTATGGG CGTGAATCAA CACATTATTG ★ GGGACAGCTA CCTGATATCG ★ TGTATGGTGG 1260  
TGGCACCGTG TTGTTTGAAG GCGTGTATGG TATTAGCGCC ATGTTGTTTG AAGGCGTGTA 1320  
TGGTATTAGC GCCATGTTGT TTGAAGGCAG CGCGTTAGTA ACAACGTTTT TTTTCGTATG 1380  
ACAACAGATA TGTATAAGAT AGCAAATTAA TGCTATTGTC ATCAATGAGA AACATAAGTA 1440  
GATATAGACC TGCCTCTTGG TAGTCCAATG ATCCATGTTG CACCTAAAAA TCCCAAAAAA 1500  
AAACATTAAG TTGTGGAGGG TAAGA

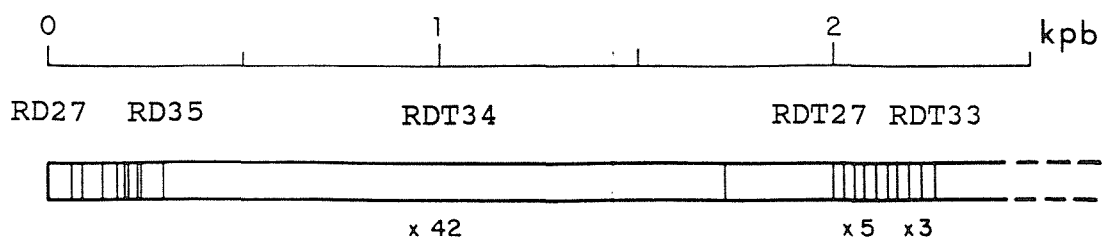
Figura 17. Secuencia determinada en el extremo derecho correspondiente a la repetición terminal invertida.

La orientación de la secuencia es la convencional, correspondiendo los últimos nucleótidos a los más próximos al extremo del genoma. Los símbolos tienen el mismo significado que en la figura 16.

**RDT34** ATTACTACTG CTGTAGGCGT TAAACATTAA AATTATTACT ACTGTTGTAG GCGTTAAACA 60  
TTAAAATTAT TACTACTGTT GTAGGCGTCA AAGATTAAAA TTATTACTAC TGCTGTAGGC 120  
GTTAAACATT AAAATTATTA CTACTGCTGT AGGCGTTAAA CATTAAACTT ATTACTACTG 180  
CTATAGGCGT TAAACATTAA AATTATTACT ACTGTTGTAG GCGTTAAACA TTAAAATTAT 240  
TACTACTGCT GTAGGCGTCA AAGATTAAAA TTATTACTAC TGCTGTAGGC GTTAAACATT 300  
AAACTTATTA CTACTGCTAT AGGCGTTAAA CATTAAAATT ATTACTACTG TTGTAGGCGT 360  
CAAAGATTAA AATTATTACT ACTGCTGTAG GCGTTAAACA TTAAAATTAT TACTACTGCT 420  
GTAGGCGTTA AACATTAAAA TTATTACTAC TGTGTAGGC GTTAAACATT AAAATTATTA 480  
CTACTGTTGT AGGCGTCAAA GATTAAAATT ATTACTACTG CTGTAGGCGT TAAACATTAA 540  
AATTATTACT ACTGCTGTAG GCGTTAAACA TTAAAATTAT TACTACTGTT GTAGGCGTTA 600  
AACATTAAAA TTATTACTAC TGCTGTAGGC GTCAAAGATT AAAATTATTA CTACTGCTGT 660  
AGGCGTTAAA CATTAAACTT ATTACTACTG CTATAGGCGT TAAACATTAA AATTATTACT 720  
ACTGTTGTAG GCGTCAAAGA TTAAAATTAT TACTACTGCT GTAGGCGTTA AACATTAAAA 780  
TTATTACTAC TGCTGTAGGC GTTAAACATT ATTACCGCTG TAGGCGTAGC TAAATTTTAA 840  
TGTTTTTTTCG TAAAAAATCG CCAGTCACAA AAAAAAGCAG CCATGTATAT CTGTGATGTT 900  
TGGTGTTTTA TCTACATGCT ATTGGCATTG CTGATATTCA TGGTAAAGTG CTATACTATA 960  
ACCATCTTAT GTGGGCTCTT TATACATCTT ATTGGCAGGA GGCCCACTAT ATATATCCTG 1020  
TTAAGTAGTT TAGCTGTATA CTATTTAATG TTTTTCGCTT GATCAAGAAA TATACAAAAA 1080  
**RDT27** TCTAAGTGCA CCATTTTATG GGCATCAATC AATCACAGCA TTCTATGGGC GTGAATCAAT 1140  
CACACCATT★ TATGGGCGTC AATCAATCAC ACCATTTIAT★ GGGCATCAAT CAATCACAGC 1200  
**RDT33** ATTCTATGGG CGTGAATCAA CACATTATTG GGGACAGCTA CCTGATATCG★ TGTATGGTGG 1260  
TGGCACCGTG TTGTTTGAAG GCGTGATGG★ TATTAGCGCC ATGTTGTTTG AAGGCGTGTA 1320  
TGGTATTAGC GCCATGTTGT TTGAAGGCAG★ CCGGTTAGTA ACAACGTTTT TTTTCGTATG 1380  
ACAACAGATA TGTATAAGAT AGCAAATTAA TGCTATTGTC ATCAATGAGA AACATAAGTA 1440  
GATATAGACC TGCCTCTTGG TAGTCCAATG ATCCATGTTG CACCTAAAAA TCCCAAAAAA 1500  
AAACATTAAG TTGTGGAGGG TAAGA

Figura 17. Secuencia determinada en el extremo derecho correspondiente a la repetición terminal invertida.

La orientación de la secuencia es la convencional, correspondiendo los últimos nucleótidos a los más próximos al extremo del genoma. Los símbolos tienen el mismo significado que en la figura 16.



RD27 61 TTTTGGCGCGC CGGAATAATT ATTTTTT 69  
201 G G ----- 228

RD35 146 CGGAATAATT ATTTTTTCGC CGGAAAGCAT TTTGT 181  
210 ----- T 245

RDT34 ATTAAAATTA TTACTACTGC TGTAGGCGTT AAAC

RDT27 CCATTTTATG GGCGTCAATC AATCACA  
■

RDT33 GTGTATGGTA TTAGCGCCAT GTTGTTTGAA GGC

Figura 18. Estructura de la repetición terminal invertida.

En la parte superior, la localización de las zonas de secuencia única y repetida. En la inferior, las secuencias de las repeticiones directas y directas en tandem. Los números indican la posición de las repeticiones directas referidas a la numeración de la figura 16. En el caso de las repeticiones directas en tandem, las posiciones subrayadas son las variables. El símbolo ■ debajo de la secuencia de RDT27 indica la posición de la base ausente en una de las repeticiones.

TABLA V. Diferencias de secuencia en la repetición terminal invertida.

K'				D'					
n°	tandem n°/posición			base	n°	tandem n°/posición			base
928	19	/	22	A	13	1	/	22	G
960	20	/	20	C	45	2	/	20	T
1008	21	/	34	C	93	3	/	34	G
1028	22	/	20	T	113	4	/	20	C

La numeración corresponde a la de las figuras 16 (K') y 17 (D'). Las RDT34 se han numerado (tandem n°) comenzando por la repetición localizada más terminalmente.

cho es más significativo en el caso de las repeticiones en tandem, de las que existe un mayor número de secuencias unidad.

En el caso de las RDT34 los tipos de secuencia existentes son 8, como se detalla en la figura 19. Si se representan por símbolos los distintos tipos de secuencias se observa a su vez un patrón de repetición. Tras las dos primeras RDT34, que son las únicas con nucleótidos distintos de las demás en las posiciones 9, 13 y 14, las 16 RDT34 siguientes se encuentran repetidas a su vez de forma directa (RDT34 número 24 a RDT34 número 39). Asimismo, el grupo de 16 RDT34 presenta a su vez subgrupos de 3 o 4 secuencias unidad repetidas, como se indica en la misma figura 19.

De las RDT34 del extremo derecho cuya secuencia se ha determinado, las cuatro situadas más terminalmente son las que contienen las diferencias de secuencia existentes entre ambas repeticiones terminales. Esta localización corresponde a las RDT34 presentes entre los dos grupos de 16 RDT34 de la repetición terminal izquierda (figura 19).

#### 4.3. Secuencia adyacente a la repetición terminal izquierda.

La secuencia de nucleótidos de la zona adyacente a la repetición terminal invertida en el extremo izquierdo se muestra en la figura 20. Se ha determinado la secuencia de 2993 nucleótidos. El análisis de dicha secuencia permitió detectar la presencia de dos repeticiones directas de 47 nucleótidos (RD47). Su localización exacta se muestra en las figuras 20 y 22 y su secuencia en la figura 23.



#### 4.4. Secuencia adyacente a la repetición terminal derecha.

De forma análoga, en la figura 21 se muestra la secuencia en la zona adyacente a la repetición terminal invertida presente en el extremo derecho. En este caso los nucleótidos cuya secuencia se ha determinado son 2752.

También en esta secuencia se detectó la presencia de repeticiones directas. Se trata en este caso de tres repeticiones directas en tandem de 25 nucleótidos (RDT25), diferenciándose la primera de ellas de las otras dos, idénticas entre sí, en solo dos posiciones. Su localización se muestra en las figuras 21 y 22 y su secuencia en la figura 23.

#### 5. SECUENCIAS CODIFICANTES.

Los experimentos de hibridación RNA-DNA y de traducción "in vitro" de mRNAs seleccionados demostraban el carácter codificante de los fragmentos terminales EcoRI, como se describe en la Introducción (I.3.2.).

Con el fin de detectar las posibles secuencias codificantes, se aplicó a la secuencia determinada en ambos extremos un programa de ordenador que localiza y representa los tripletes de iniciación y terminación en todas las fases de lectura.

Los resultados obtenidos al representar la posición de fases de lectura abierta de un tamaño superior a 300 nucleótidos se muestran en la figura 22. No se representa la repe-

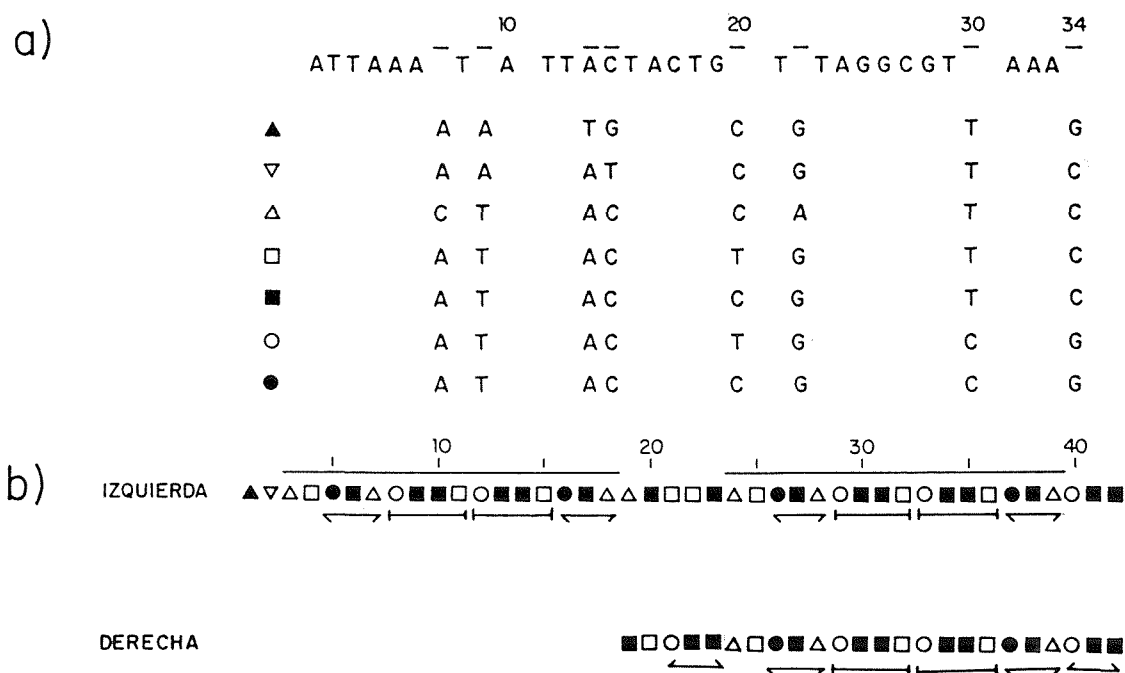


Figura 19. Microheterogeneidades dentro de las RDT34.

En la parte a) se asigna un símbolo a cada una de las combinaciones de bases encontradas en las ocho posiciones variables. En la parte b) se muestra por medio de dichos símbolos la secuencia de microheterogeneidades encontrada en ambos extremos, correspondiendo la orientación izquierda-derecha a la terminal-interna en el genoma. En el caso del extremo derecho, el primer símbolo indica la primera repetición presente de forma incompleta en el fragmento terminal clonado. Las líneas superiores señalan el grupo de 16 repeticiones, y las inferiores los subgrupos de 3 o 4 unidades.

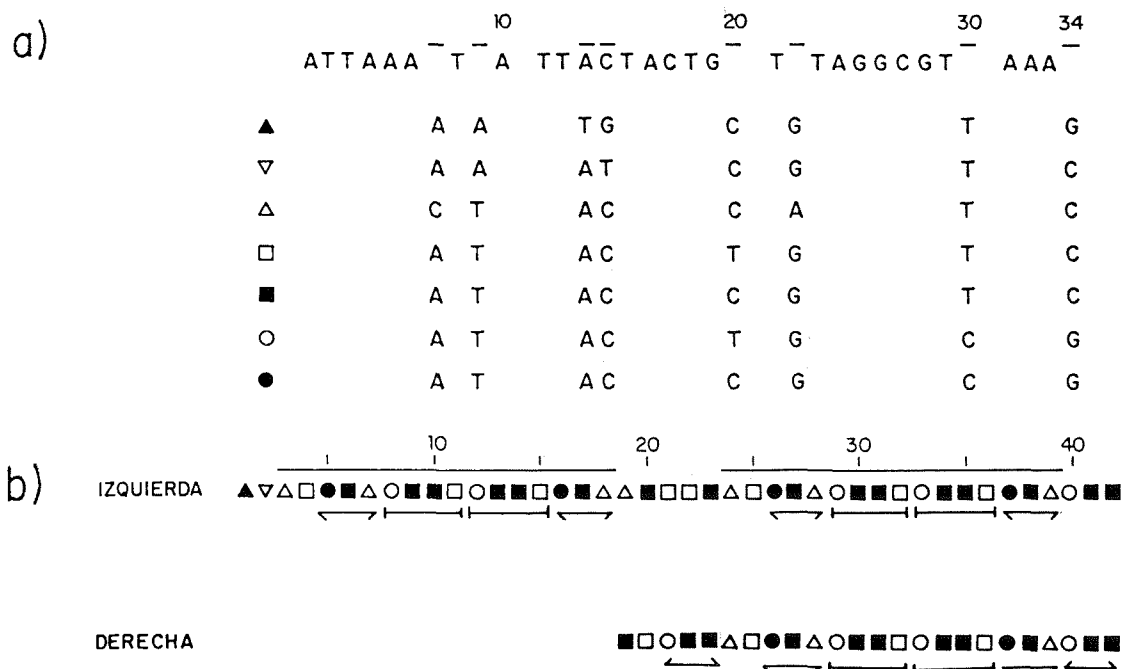





Figura 19. Microheterogeneidades dentro de las RDT34.

En la parte a) se asigna un símbolo a cada una de las combinaciones de bases encontradas en las ocho posiciones variables. En la parte b) se muestra por medio de dichos símbolos la secuencia de microheterogeneidades encontrada en ambos extremos, correspondiendo la orientación izquierda-derecha a la terminal-interna en el genoma. En el caso del extremo derecho, el primer símbolo indica la primera repetición presente de forma incompleta en el fragmento terminal clonado. Las líneas superiores señalan el grupo de 16 repeticiones, y las inferiores los subgrupos de 3 o 4 unidades.

ATAACCTAGC TAGCATTTCG TTCAATGTTA TTTCATTCAA TATAAAGTTG ATCGATATCT 1800  
TCTCTAGAAA ACAACAAATT ATTATTTTTA ATTCTCTAT ATTCTGGAAA AGGGATTAT 1860  
TAGATAACAA TTTATGGCAT AGAATAATAT TACTACTAGT TTTAATACGA TGTATTTTAT 1920  
AAAATATTTG TACAATATCC ATTTCATTCA AAATTTTTCG GCCTAACTCC CGGCAGAAAT 1980  
TCCAAGTAG CTCCGTATTG ACAGTGACTA AGCTAGAGTT GATGCTGCA CCCCATTCGG 2040  
TAAACAACCTC TATTAGATCA TAGTTGTTTT CCTGCACAGC TTTCATTAAAT GCGAGATTTA 2100  
ACTCTAAACC ATCTTTAAAA ATTGCTGATT TTATCATCAA TTGATTATCC TCATTAGTAG 2160  
AAAACATAAT TGGAGCTCCA TGCCACCACA AACCAACAATA TTTCAAAATA AAGTAGTGAT 2220  
TTTTTGGATAT GTGCTGTGTG GCCAGTATTT TTTTAGCAAG AGCCTGCAGA GAAAGTGGAG 2280  
TAGACATTTT  TTTTTGCAAA ATGCTTTAAG TTTTAATGTT TTCAAGAATA CAGATTGGAT 2340  
AAATTAGGTT GACATTAGTA ACAGAAAAGTA TTCAATATTA TGTAGACCTA TAAAAAATGA 2400  
K'362 GACCTCCAA AAAATAATAA ACAACACAAA AAAAATATGT TTAATATTA AATGACAATT 2460  
TCTACATTGC TTATTGCTCT TATTATACTA GTTATTATTA TTTTAGTCGT ATTTTPTAT 2520  
TATAAAAAAC AACAAACCACC AAAAAAGGTC TGTAAAGTAG ATAAAGATTG TGGTAGTGA 2580  
GAGCATTTGT TTCGTGGAAC ATGCAGCACA TTGAGCTGCT TAGACGCTGT AAAAAATGGAC 2640  
AAACGAAATA TTAAGATAGA TTCTAAGATT TCCTCATGIG AATTCACCTCC CAATTTTAC 2700  
CGTTTTACGG ATACTGCTGC CGATGAGCAG CAAGAATTG GAAAAACACG GCATCCTATA 2760  
AAAATAACTC CATCTCCAAG TGAATCCCAT AGCCCCAAG AGGTGTGTA AAAATATTGT 2820  
TCATGGGGAA CCGATGACTG TACAGGTTGG GAATATGTTG GTGATGAAAA GGAGGGAACA 2880  
TGTATGTAT ATAATAATCC ACATCACCCG GTTCTTAAAT ATGGTAAGGA TCACATCATA 2940  
GCCTTACCTA GAAATCATAA ACATGCATAA ATAAATACAT TAGGCTCATC GTA

K'L177

Figura 20. Secuencia de nucleótidos de la zona adyacente a la repetición terminal invertida izquierda.

La orientación es la convencional, correspondiendo los primeros nucleótidos a los más próximos a la RTI. Se señalan simbólicamente los codones de iniciación () y terminación () de las fases de lectura abierta y la posición de RD47 (Δ).

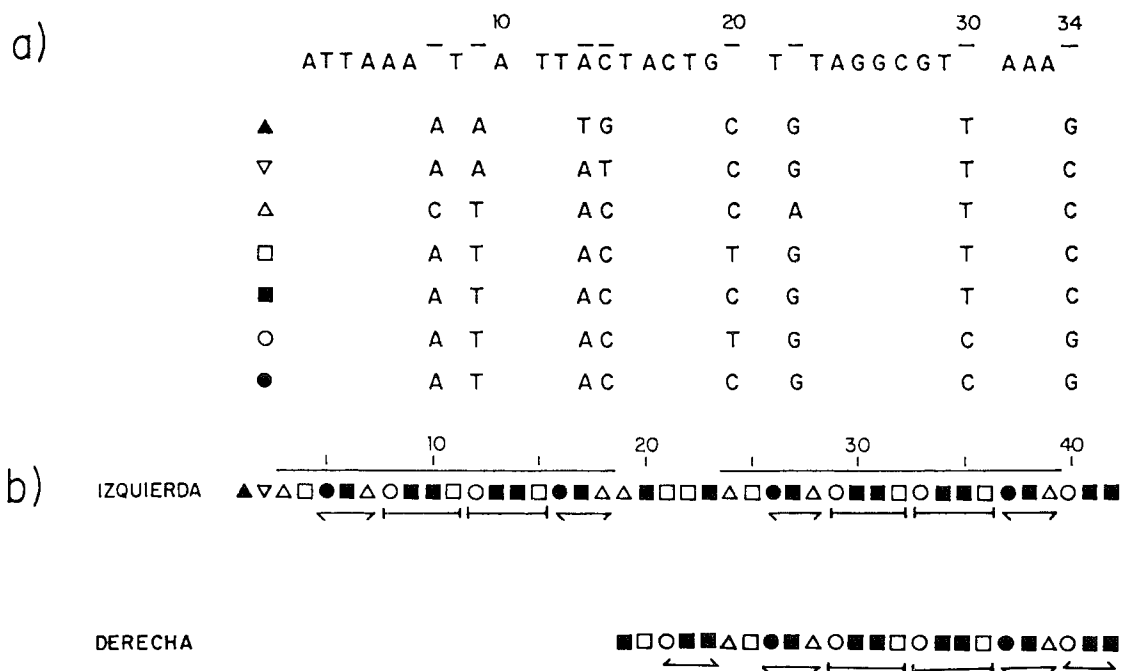


Figura 19. Microheterogeneidades dentro de las RDT34.

En la parte a) se asigna un símbolo a cada una de las combinaciones de bases encontradas en las ocho posiciones variables. En la parte b) se muestra por medio de dichos símbolos la secuencia de microheterogeneidades encontrada en ambos extremos, correspondiendo la orientación izquierda-derecha a la terminal-interna en el genoma. En el caso del extremo derecho, el primer símbolo indica la primera repetición presente de forma incompleta en el fragmento terminal clonado. Las líneas superiores señalan el grupo de 16 repeticiones, y las inferiores los subgrupos de 3 o 4 unidades.





ATAACCTAGC TAGCATTTTCG TTCAATGTTA TTTTCATTCAA TATAAAGTTG ATCGATATCT 1800  
 TCTCTAGAAA ACAACAAAATT ATTATTTTTA ATTCTCTCTAT ATTCTGGAAA AGGGGATTAT 1860  
 TAGATAACAA TTTATGGCAT AGAATAATAT TACTACTAGT TTTAATAACGA TGTATTTTTAT 1920  
 AAAATATTTG TACAATATCC ATTTCATTCA AAATTTTTCG GCCTAACTCC CGGCAGAAAT 1980  
 TCCAAGTATG CTCCGTATTG ACAGTGACTA AGCTAGAGTT GATGTCTGCA CCCCATTCGG 2040  
 TAAACAACCTC TATTAGATCA TAGTTGTTTT CCTGCACAGC TTTTCATTAAAT GCGAGATTTA 2100  
 ACTCTAAACC ATCTTTTAAAA ATTGCTGATT TTATCATCAA TTGATTATCC TCATTAGTAG 2160  
 AAAACATAAT TGGAGCTCCA TGCCACCACA AACCAACAATA TTTCAAAAATA AAGTAGTGAT 2220  
 TTTTGGATAT GTGCTGTGTG GCCAGTATTT TTTTAGCAAG AGCCTGCAGA GAAAGTGGAG 2280  
 TAGACATTTT  TTTTTCGAAA ATGCTTTAAG TTTTAAATGTT TTCAAGAATA CAGATTGGAT 2340  
 AAATTAGGTT GACATTAGTA ACAGAAAAGTA TTCAATATTA TGTAGACCTA TAAAAAATGA 2400  
 K'362 GACCCCTCAA AAAATAATAA ACAACACAAA AAAAATATGT TTAATATTAA AATGACAATT 2460  
 TCTACATTGC TTATTGCTCT TATTTACTA GTTATTATTA TTTTAGTCGT ATTTTATAT 2520  
 TATAAAAAAC AACAAACCACC AAAAAAGGTC TGTAAAGTAG ATAAAGATTG TGGTAGTGA 2580  
 GAGCATTTGT TTCGTGGAAC ATGCAGCACA TTGAGCTGCT TAGACGCTGT AAAAAATGGAC 2640  
 AACGGAATA TTAAGATAGA TTCTAAGATT TCCTCATGTG AATTCACTCC CAATTTTAC 2700  
 CGTTTTACGG ATACTGCTGC CGATGAGCAG CAAGAATTG GAAAAACACG GCATCCTATA 2760  
 AAAATAACTC CATCTCCAAG TGAATCCCAT AGCCCCAAG AGGTGTGTGA AAAATATTGT 2820  
 TCATGGGGAA CCGATGACTG TACAGGTTGG GAATATGTTG GTGATGAAAA GGAGGGAACA 2880  
 TGTATGTAT ATAATAATCC ACATCACCCG GTTCTTAAAT ATGCTAAGGA TCACATCATA 2940  
 GCCTTACCTA GAAATCATAA ACATGCATAA ATAAATACAT TAGGCTCATC GTA

Figura 20. Secuencia de nucleótidos de la zona adyacente a  
 la repetición terminal invertida izquierda.

La orientación es la convencional, correspondiendo los  
 primeros nucleótidos a los más próximos a la RTI. Se señalan  
 simbólicamente los codones de iniciación () y terminación  
 (  ) de las fases de lectura abierta y la posición de RD47 (  ).

	GGATCCCTAA	TACGCTCCAT	CCAGAAGAAT	AATTAAAAAA	AATATTTTTT	TTGGTAAGTT	60
	TATAAACTAT	ATAGTTAAAT	ATGGTAAAAA	AAATCACATA	ATAATTAAAC	TGAACGTGTT	120
	AGAATTAATA	TTTTTTTATA	ATCGGATATA	ATATCCATTA	AATCAATAAA	TGATAGTGTT	180
	GCTACCACAC	TAAACAATAA	CAAACAGAAA	CGCATGATAC	CTTTCCTCAT	GATTTATAAT	240
	AGCGAGTTAT	CTAAAGATTT	TTTTGAAAAA	AATATTAAAT	TTTAGTTGAT	ATATTTTTTT	300
	AGTTACAACA	TTGCTTTAGA	AAAAAACCT	AATTACTACA	TAGCAAATAA	AGCGAGCGTG	360
	TTGTTACAAA	CAACATTTTT	TTGCGCCTGG	ATACTCTTAT	ATATGAGAAC	TATAATATGC	420
	CATATTAAATC	CTATTACCAA	CAATGCCAAA	AATAGTATGT	AGGCAATGAC	AACTTTAAAT	480
	ACAAAATTTT	CATGGTTATC	TCTAAAAATC	TTGAAAAAAT	ATTTAATTTT	AGATCGGTCA	540
	CCTACGGTAG	TAATACTAAT	TTTAATAATT	GATGACTAAA	ATCATATTAT	AATGCCGTGC	600
<b>D'363</b>	AAAAAATAAT	TATTTTTCGG	TTAAAGGATA	CCTTTAAATA	AAAAACGATG	CCATCCACTC	660
	TACAAGTGCT	TGCTAAAAAG	GTGTTGGCCT	TAGGGGAACA	TAAAGAAAAT	GAACATATAT	720
	CTAGAGAATA	TTATTATCAT	ATATTAAAGT	GTTGCGGTTT	ATGGTGGCAT	GAGGCTCCGA	780
	TTATACTTTG	TTTTGATGGG	AGTGAGCAAA	TGATGATAAA	GACTCCAATC	TTTGAAGAAG	840
	GCATATTACT	TAATACTGCA	TTAATGAAAG	CTGTACAGGA	GAACAATTAT	GAATTAATAA	900
	ACTTGTTTAC	TGAATGGGGA	GCAAACATCA	ATTATGGATT	AATTTCCATT	AATACTGAGC	960
	ATGCCCCGGA	CCTATGTCGA	AAATTAGGAG	CTAAAGAAAT	GCTTGAAAGA	AATGAAGTTA	1020
	TACAAATTGT	ATTTAAAACA	TTAGATGATA	TCACCAGTAG	TAATATAATT	TTATGTCATG	1080
	AATTATTTAC	CAACAATCCT	CTTTTAGAGA	ATGTAAATAT	GGGGGAAATG	AGGATGATAA	1140
	TTCATTGGAG	GATGAAAAAT	TTAACGAACC	TATTATTAAA	TAATGACTCT	ATTAGTGAAA	1200
	TATTAECTAA	ATTCTGGTAT	GGTATAGCAG	TAAAATATAA	TCTTAAGGAT	GCAATCCAAT	1260
	ATTTTACCA	GAGATTCATG	GACTTCAACG	AGTGGCGAGT	AACATGTGCT	CTTCTTTT	1320
	ATAATGTGAA	TGATCTTCAT	AAGATGTATA	TAACAGAGAA	GGTTCATACG	AATAATGACG	1380
	AAATGATGAA	TCTAGCCTGC	AGCATTCAAG	ACAGAAATTT	ATCAACCATT	TACTATTGTT	1440
	TTCTATTGGG	GGCTAACATC	AATCAAGCAA	TGTTGACCTC	AGTATTAAAT	TATAATATTT	1500
	TTAACTTATT	CTTTTGTATA	GACTTAGGGG	CTGATGCCTT	TGAAGAGGGT	AAGACCTGG	1560
	CGAAACAAAA	GGGGTATAAT	GAAATAGTGG	AAATCTTATC	ATTAGATATC	ATTTATAGTC	1620
	CAAATACTGA	CTTCTCATCA	AAAATAGAAC	CTGAACATAT	TAGTTCTTTG	TTAAAAAACT	1680
	TTTATCCAAA	AAATCTGTTC	GCTTTTGATC	GTTGCAATCC	TGGTTTATAT	TATTCTTAGA	1740
	GGACCGCTAC	AAAAATTATT	TTTTTTTCTT	GATCAAAGCT	CCAAAATAAT	TATTAGATTA	1800
	AAGTCGCTA	TAGCAGCTGC	CCACTCCAAA	AAAAGTATTT	TATAGTACAA	AAAACACGAA	1860
	AAATAATTTG	CGGCCGGCGG	CAACTATTTG	TTGTTGTCCA	AACTTAAATG	TTTTTTTAAAT	1920
	ATTTTAAAT	GCAACCATGG	ATTGTTGGAC	TATCAGGGAG	AAGAACTATA	GCTACATCAT	1980
	ATTGTCAATA	CTGGTAATAC	TATTAATATG	GTATCTTATA	CTTAACATAT	GTGATCGAA	2040
	AAAAAATGCA	GTTACAAACA	ACATGCCGCC	ACCAGCGTAC	ACGGTGTCAA	GTAGCTGTTT	2100
	TCAATAATAG	GGTTGATCGA	CGCTCTTCGT	AATAATATGT	TGATTGGCGC	ATCATAAAAT	2160
	GCTGAGGAAC	TATCCGTTAA	TATATGTTGA	TTGACGCCCA	CTTATGAATG	GAAGTAATGA	2220
	TTTTGTATAA	AATACGGGTT	TGTGAGGGCC	TTATTTTTTT	CTTATTAGAA	CAAAGCGTGT	2280
	ATTTTAAAGG	CCTATAGCAA	GAGTATGTTT	AATTAACACC	TACAACAGTA	ATATTTAAGG	2340
	CCAGTAAAAT	AATGTTTAAT	TAAGGCCTGA	CCACTAAAAC	TTAAACGATT	TTGTAAAAAA	2400
<b>D'42</b>	ATATGCCTAC	TCCACTTTCT	CTACAGGCTC	CCGCTAAAAA	AGTACTGGCC	ACACAGCACA	2460
	TATCTAAAGA	TCACCTTTAT	TTTGAAATAT	TGTGGTTTAT	GGTGGCATT	TTTGACGCCT	2520
<b>RDT25</b>	ACAGTCTTTA	ACGCCTGCAG	TAATAATTGA	TATCTCCAGC	GCTAATAATA	ATTGATACCC	2580
	ACAGCGGTAA	TAATAATTGA	TACCCACAGC	GGTAATGTCG	TCCATCTGGC	CGCCCCAAAA	2640
	AAAAGTATTT	ACGGTGGGGT	TTATTACCGG	CGGCGTAACA	CCAGTTATGG	TTAGTTTTGT	2700
	CTGGCCCGCC	GCCCAGCCGC	AAAAAAAAT	CAATTACAGC	CGCAAAAAAA	AATATTTCCG	2760
	GCCGCGGAGT	TTCTACAAAA	AAAATGTTAG	CTTTTAATTG	TTGTAGACCA	GGTTCTTAGG	2820
	GAAGGTTCAA	CTCCTTTTTC	TCGAATAAAA	AA			

Figura 21. Secuencia de nucleótidos de la zona adyacente a la repetición terminal invertida derecha.

La orientación es la convencional, correspondiendo los últimos nucleótidos a los más próximos a la RTI. Se señalan simbólicamente los codones de iniciación y terminación de las fases de lectura abierta como en la figura 20. Los símbolos denotan la posición de las RDT25.

	GGATCCCTAA	TACGCTCCAT	CCAGAAGAAT	AATTAAAAAA	AATATTTTTT	TTGGTAAGTT	60
	TATAAACTAT	ATAGTTAAAT	ATGGTAAAAA	AAATCACATA	ATAATTAAAC	TGAACGTGTT	120
	AGAATTAATA	TTTTTTTATA	ATCGGATATA	ATATCCATTA	AATCAATAAA	TGATAGTGTT	180
	GCTACCACAC	TAAACAATAA	CAAACAGAAA	CGCATGATAC	CTTTCCTCAT	GATTTATAAT	240
	AGCGAGTTAT	CTAAAGATTT	TTTTGAAAAA	AATATTAAAT	TTTAGTTGAT	ATATTTTTTT	300
	AGTTACAACA	TTGCTTTAGA	AAAAAACCT	AATTACTACA	TAGCAAATAA	AGCGAGCGTG	360
	TTGTTACAAA	CAACATTTTT	TTGCGCCTGG	ATACTCTTAT	ATATGAGAAC	TATAATATGC	420
	CATATTAATC	CTATTACCAA	CAATGCCAAA	AATAGTATGT	AGGCAATGAC	AACTTTAAAT	480
	ACAAAATTTT	CATGGTTATC	TCTAAAAATC	TTGAAAAAAT	ATTTAATTTT	AGATCGGTCA	540
	CCTACGGTAG	TAATACTAAT	TTTAATAATT	GATGACTAAA	ATCATATTAT	AATGCCGTGC	600
D'363	AAAAAATAAT	TATTTTTTCGG	TTAAAGGATA	CCTTTAAATA	AAAAACGATG	CCATCCACTC	660
	TACAAGTGCT	TGCTAAAAAG	GTGTTGGCCT	TAGGGGAACA	TAAAGAAAAT	GAACATATAT	720
	CTAGAGAATA	TTATTATCAT	ATATTAAAGT	GTGCGGTTT	ATGGTGGCAT	GAGGCTCCGA	780
	TTATACTTTG	TTTTGATGGG	AGTGAGCAAA	TGATGATAAA	GACTCCAATC	TTTGAAGAAG	840
	GCATATTACT	TAATACTGCA	TTAATGAAAG	CTGTACAGGA	GAACAATTAT	GAATTAATAA	900
	ACTTGTTTAC	TGAATGGGGA	GCAAACATCA	ATTATGGATT	AATTTCCATT	AATACTGAGC	960
	ATGCCCCGGA	CCTATGTCGA	AAATTAGGAG	CTAAAGAAAT	GCTTGAAAGA	AATGAAGTTA	1020
	TACAAATTGT	ATTTAAAACA	TTAGATGATA	TCACCAGTAG	TAATATAATT	TTATGTCATG	1080
	AATTATTTAC	CAACAATCCT	CTTTTAGAGA	ATGTAAATAT	GGGGGAAATG	AGGATGATAA	1140
	TTCATTGGAG	GATGAAAAAT	TTAACGAACC	TATTATTAAA	TAATGACTCT	ATTAGTGAAG	1200
	TATTAACATA	ATTCTGGTAT	GCTATAGCAG	TAAAATATAA	TCTTAAGGAT	GCAATCCAAT	1260
	ATTTTACCA	GAGATTCATG	GACTTCAACG	AGTGGCGAGT	AACATGTGCT	CTTCTTTTAA	1320
	ATAATGTGAA	TGATCTTCAT	AAGATGTATA	TAACAGAGAA	GGTTCATACG	AATAATGACG	1380
	AAATGATGAA	TCTAGCCTGC	AGCATTCAAG	ACAGAAATTT	ATCAACCATT	TACTATTGTT	1440
	TTCTATTGGG	GGCTAACATC	AATCAAGCAA	TGTTGACCTC	AGTATTAAAT	TATAATATTT	1500
	TTAACTTATT	CTTTTGTATA	GACTTAGGGG	CTGATGCCTT	TGAAGAGGGT	AAGACCCTGG	1560
	CGAAACAAAA	GGGGTATAAT	GAAATAGTGG	AAATCTTATC	ATTAGATATC	ATTTATAGTC	1620
	CAAATACTGA	CTTCTCATCA	AAAATAGAAC	CTGAACATAT	TAGTTCTTTG	TTAAAAAACT	1680
	TTTATCCAAA	AAATCTGTTC	GCTTTTGATC	GTTGCAATCC	TGGTTTATAT	TATTCTTAGA	1740
	GGACCGCTAC	AAAAATTATT	TTTTTTTCTT	GATCAAAGCT	CCAAAATAAT	TATTAGATTA	1800
	AAGTCGCCTA	TAGCAGCTGC	CCACTCCAAA	AAAAGTATTT	TATAGTACAA	AAAACACGAA	1860
	AAATAATTTG	CGGCCGGCGG	CAACTATTTG	TTGTTGTCCA	AAACTTAATG	TTTTTTTAAAT	1920
	ATTTTTTAAAT	GCAACCATGG	ATTGTTGGAC	TATCAGGGAG	AAGAACTATA	GCTACATCAT	1980
	ATTGTCAATA	CTGGTAATAC	TATTAATATG	GTATCTTATA	CTTAACATAT	GTCGATCGAA	2040
	AAAAAATGCA	GTTACAAACA	ACATGCCCGC	ACCAGCGTAC	ACGGTGTCAA	GTAGCTGTTC	2100
	TCAATAATAG	GGTTGATCGA	CGCTCTTCGT	AATAATATGT	TGATTGGCGC	ATCATAAAAT	2160
	GCTGAGGAAC	TATCCGTTAA	TATATGTTGA	TTGACGCCCA	CTTATGAATG	GAAGTAATGA	2220
	TTTTGTATAA	AATACGGGTT	TGTGAGGGCC	TTATTTTTTT	CTTATTAGAA	CAAAGCGTGT	2280
	ATTTTTAAGG	CCTATAGCAA	GAGTATGTTT	AATTAACACC	TACAACAGTA	ATATTTAAGG	2340
	CCAGTAAAT	AATGTTTAAT	TAAGGCCTGA	CCACTAAAAC	TTAAACGATT	TTGTAAAAAA	2400
D'42	ATATGCCTAC	TCCACTTTCT	CTACAGGCTC	CCGCTAAAAA	AGTACTGGCC	ACACAGCACA	2460
	TATCTAAAGA	TCACCTTTAT	TTTGAAATAT	TGTGCTTTAT	GGTGGCATTC	TTTGACGCCT	2520
RDT25	ACAGTCTTTA	ACGCCTGCAG	TAATAATTGA	TATCTCCAGC	GCTAATAATA	ATTGATACCC	2580
	ACAGCGGTAA	TAATAATTGA	TACCCACAGC	GGTAATGTCT	TCCATCTGGC	CGCCCCAAAA	2640
	AAAAGTATTT	ACGGTGGGGT	TTATTACCGG	CGGCGTAACA	CCAGTTATGG	TTAGTTTGTG	2700
	CTGGCCCGCC	GCCCAGCCGC	AAAAAAAAT	CAATTACAGC	CGCAAAAAAA	AATATTTCCG	2760
	GCCGCGGAGT	TTCTACAAAA	AAAATGTTAG	CTTTTAATTG	TTGTAGACCA	GGTTCTTAGG	2820
	GAAGGTTCAA	CTCCTTTTTT	TCGAATAAAA	AA			

Figura 21. Secuencia de nucleótidos de la zona adyacente a la repetición terminal invertida derecha.

La orientación es la convencional, correspondiendo los últimos nucleótidos a los más próximos a la RTI. Se señalan simbólicamente los codones de iniciación y terminación de las fases de lectura abierta como en la figura 20. Los símbolos denotan la posición de las RDT25.



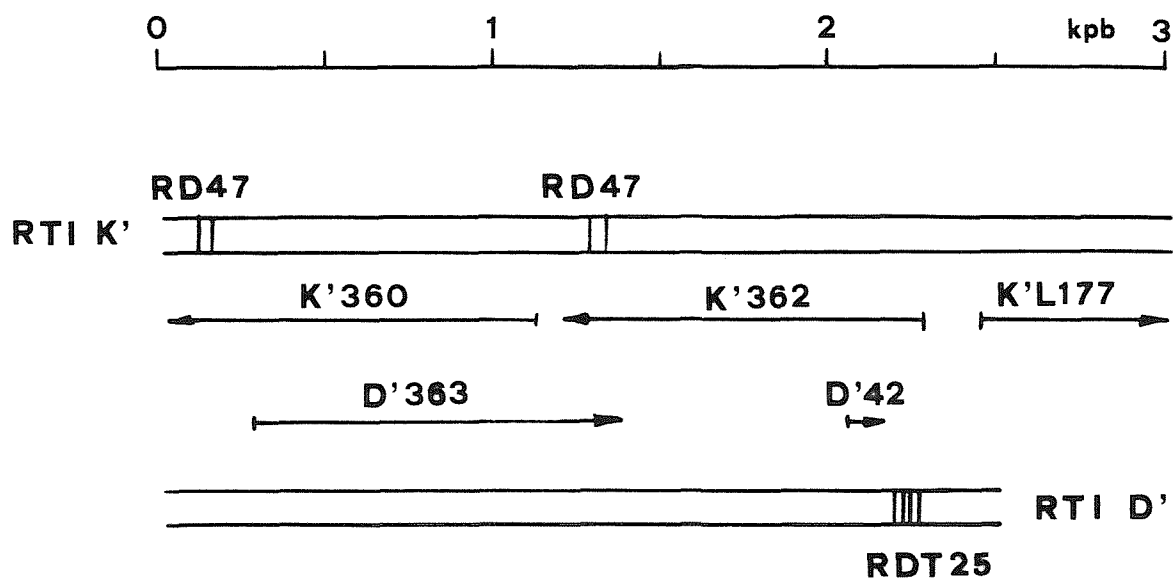


Figura 22. Localización de las fases de lectura abierta y repeticiones directas en las secuencias adyacentes a la repetición termina invertida.

Se indica la longitud de la fase en codones. El inicio de las flechas corresponde al codón de iniciación. Las flechas apuntando hacia la izquierda indican que la secuencia supuestamente codificante corresponde a la cadena inferior, y viceversa.

RD47

115  
A A G G C G T T A A T T T T T T C C G G A T C T G  
1284  
A A G G C A T T A A T T T T T T C T G G A T C T G

T C G T T T T T A T T G A T A A A A G A G A 161  
T C G T T T T T A A A G A T A A A A G A G A 1330

RDT25

2541  
T A A T A A T T G A T A T C T C C A G C G C T A A  
T A A T A A T T G A T A C C C A C A G C G G T A A  
T A A T A A T T G A T A C C C A C A G C G G T A A 2615

Figura 23. Secuencias de las repeticiones directas presentes en la zona adyacente a la RTI (RD47 y RDT25).

Las posiciones variables se muestran subrayadas. La numeración corresponde a las de las figuras 20 y 21.

tición terminal invertida debido a la ausencia en ella de fases de lectura abierta del tamaño anteriormente indicado. La nomenclatura empleada indica el fragmento EcoRI en que se encuentra la fase, así como su número de codones. La posición exacta de los codones de iniciación o terminación de dichas fases se muestra asimismo en las figuras 20 y 21.

La secuencia adyacente en posición 5' al ATG iniciador de todas las fases contiene zonas homologas a la secuencia consenso de unión del RNA mensajero al ribosoma eucariotico. Dichas zonas se muestran, subrayadas, en la figura 24.

Las fases K'360, K'362 y D'363 presentan un elevado grado de homología de secuencia, tanto de nucleótidos como de aminoácidos. La fase D'42 consta solo de 42 codones, de los cuales los 30 primeros muestran un alto grado de homología con el comienzo de las tres fases anteriormente citadas. Por esta razón se analizan a continuación estas cuatro fases de forma conjunta, denominándoselas familia 360.

#### 5.1. Familia 360.

Las secuencias de nucleótidos de las fases K'360, K'362, D'363 y D'42 y su traducción a aminoácidos se muestran, alineadas, en la figura 25. En dicha figura se señalan los aminoácidos idénticos y homólogos localizados en posiciones análogas de las secuencias.

La composición de aminoácidos de las proteínas que se sintetizarían si las fases K'360, K'362 y D'363 se traduje-

sen de forma completa se muestra en la tabla VI. Los puntos isoeléctricos teóricos de cada proteína son distintos y se muestran asimismo en la tabla VI.

La cuantificación de las diferencias existentes entre las secuencias, tanto de nucleótidos como de aminoácidos, de las cuatro fases, se muestra en la figura 26a. Esta divergencia de secuencia se puede representar gráficamente en forma de dendrograma (figura 26b). Si se consideran las secuencias completas de K'360, K'362 y D'363, el par que presenta un menor número de diferencias es K'360 y K'362. Al comparar solamente los 30 primeros triplete de las cuatro fases, las que presentan un menor número de diferencias entre si son D'42 y K'362.

La frecuencia de uso de codones deducida a partir de las secuencias de la familia 360 se muestra en la tabla VII. Se observa un uso de codones muy similar en las fases K'360, K'362 y D'363.

Los residuos idénticos en posiciones análogas de las secuencias se agrupan preferencialmente en ciertas zonas de las mismas, según se observa en la figura 25. Se han dividido las secuencias en segmentos intentando maximizar la separación entre las zonas con mayor y menor homología de secuencia. La cuantificación de la identidad, tanto de nucleótidos como de aminoácidos, en cada segmento se muestra en la tabla VIII.

En todas las comparaciones, los segmentos I, III, V, VII y X son los que presentan el mayor grado de identidad,

CTAAAACTTAAACGATTTTGTAAAAAATATG	D' 42
TTAAAACTTAAACCATTTTGCAAAAAAATG	K'362
TTCGGAAATAATTATTTTGCAGATAGAAGATG	K'360
GGTTAAAGGATACCTTTAAATAAAAAACGATG	D'363
<u>CCCTCCAAAAAATAATAACAACACAAAAAATATG</u>	K'L177

GATCPPPPPGPAGGTTTACC

Figura 24. Secuencias adyacentes al ATG iniciador en las distintas fases de lectura abierta.

La última línea corresponde a la secuencia complementaria a la del RNA ribosomal 18S implicada en la unión de RNAs mensajeros (subrayada)(170). Las zonas homólogas a dicha secuencia presentes en la zona adyacente al iniciador se muestran asimismo subrayadas. P= pirimidina.

TABLA VI. Composición de aminoácidos de las proteínas co-  
rrespondientes a las fases K'362, K'360 y D'363.

	K'362	K'360	D'363
A	20	20	18
C	10	13	10
D	14	19	15
E	21	24	28
F	22	17	19
G	8	12	14
H	11	9	10
I	39	33	34
K	27	22	21
L	45	45	44
M	13	14	14
N	33	28	37
P	5	7	8
Q	9	12	9
R	13	6	10
S	22	20	19
T	15	22	17
V	12	12	12
W	7	6	6
Y	16	19	18
Pm	42638	41692	42322
pI	9.25	4.85	5.25

Los aminoácidos se denominan con el código de una letra.  
Se muestran el peso molecular y el punto isoeléctrico teórico (Pm y pI).

Figura 25. Comparación de las secuencias de nucleótidos de las fases de lectura abierta D'42, K'362, K'360 y D'363 y su traducción a aminoácidos. Los guiones indican interrupciones introducidas para maximizar la homología. El símbolo indica identidad de aminoácido en posiciones análogas de las secuencias de K'362, K'360 y D'363. El símbolo o indica aminoácidos homólogos según los grupos: Ser, Thr, Pro, Ala, Gly; Met, Ile, Leu, Val; Asp, Glu, Asn, Gln; Arg, Lys, His; Phe, Tyr, Trp (172). Los límites entre segmentos con distinto grado de homología ( en números romanos) se denotan con trazos verticales.

D42 1 ATG CCT ACT CCA CTT TCT CTA CAG GCT CCC GCT AAA AAA GTA CTG CCC ACA --- --- CAG  
Met Pro Thr Pro Leu Ser Leu Gln Ala Pro Ala Lys Lys Val Leu Ala Thr Gln

K362 1 ATG TCT ACT CCA CTT TCT CTG CAG GCT CTT GCT AAA AAA ATA CTG CCC ACA --- --- CAG  
Met Ser Thr Pro Leu Ser Leu Gln Ala Leu Ala Lys Lys Ile Leu Ala Thr Gln

K360 1 ATG --- --- CCA TCT ACT TTA CAA GCA CTT ACT AAA AAA GTA CTA GCT ACA --- --- CAG  
Met Pro Ser Thr Leu Gln Ala Leu Thr Lys Lys Val Leu Ala Thr Gln

D363 1 ATG --- --- CCA TCC ACT CTA CAA GTG CTT GCT AAA AAG GTG TTG GCC TTA GGG GAA CAT  
Met Pro Ser Thr Leu Gln Val Leu Ala Lys Lys Val Leu Ala Leu Gly Glu His

I  
61 --- --- --- CAC ATA TCT AAA GAT CAC CTT TAT TTT GAA ATA TTG TGG TTT ATG GTG  
His Ile Ser Lys Asp His Leu Tyr Phe Glu Ile Leu Trp Phe Met Val

61 --- --- --- CAC ATA TCC AAA AAT CAC --- TAC TTT --- ATT TTG AAA TAT TGT GGT  
His Ile Ser Lys Asn His Tyr Phe Ile Leu Lys Tyr Cys Gly

61 --- --- --- CCT GTG TTC AAA GAT CAT --- TAT TGT --- ATT TTA GAG CGT TGT GGT  
Pro Val Phe Lys Asp Asp Tyr Cys Ile Leu Glu Arg Cys Gly

61 AAA GAA AAT GAA CAT ATA TCT AGA GAA TAT --- TAT TAT CAT ATA TTA AAG TGT TGC GGT  
Lys Glu Asn Glu His Ile Ser Arg Glu Tyr Tyr Tyr His Ile Leu Lys Cys Cys Gly

III  
121 GCA TTC TTT GAC GCC TAC AGT CTT  
Ala Phe Phe Asp Ala Tyr Ser Leu

121 TTG TGG TCG CAT CGA GCT CCA ATT ATG TTT TCT --- ACT AAT GAG GAT AAT CAA TTG ATG  
Leu Trp Trp His Gly Ala Pro Ile Met Phe Ser Thr Asn Glu Asp Asn Gln Leu Met

121 TTG TGG TGG CAT GAA GCC CCA ATT ACG ATT CAT CAT ACT TGT ATA GAT AAA CAA ATA TTA  
Leu Trp Trp His Glu Ala Pro Ile Thr Ile His His Thr Cys Ile Asp Lys Gln Ile Leu

121 TTA TGG TCG CAT GAG GCT CCG ATT ATA CTT TGT --- TTT GAT GGG AGT GAG CAA ATG ATG  
Leu Trp Trp His Glu Ala Pro Ile Ile Leu Cys Phe Asp Gly Ser Glu Gln Met Met

IV  
181 ATA AAA TCA GCA ATT TTT AAA GAT GGT TTA GAG TTA AAT CTC GCA TTA ATG AAA GCT GTG  
Ile Lys Ser Ala Ile Phe Lys Asp Gly Leu Glu Leu Asn Leu Ala Leu Met Lys Ala Val

181 ATA AAA ACA GCA AGT TTT AAA CAT GGT TTA ACA TTA AAT GTT GCA TTA ATG AAG GCT GTA  
Ile Lys Thr Ala Ser Phe Lys His Gly Leu Thr Leu Asn Val Ala Leu Met Lys Ala Val

181 ATA AAG ACT CCA ATC TTT GAA GAA GGC ATA TTA CTT AAT ACT CCA TTA ATG AAA GCT GTA  
Ile Lys Thr Pro Ile Phe Glu Glu Gly Ile Leu Leu Asn Thr Glu Trp Gly Ala Asp Ile Ser Phe

V  
241 CAG GAA AAC AAC TAT GAT CTA ATA GAG TTG TTT ACC GAA TGG GGT GCA GAC ATC AAC TCT  
Gln Glu Asn Asn Tyr Asp Leu Ile Glu Leu Phe Thr Glu Trp Gly Ala Asp Ile Asn Ser

241 CAG GAA AAC AAT CAT GGT TTA ATA GAG CTG TTT ACC GAA TGG GGT GCA GAC ATC AGC TTT  
Gln Glu Asn Asn His Gly Leu Ile Glu Leu Phe Thr Glu Trp Gly Ala Asp Ile Ser Phe

241 CAG GAG AAC AAT TAT GAA TTA ATA AAC TTG TTC ACT GAA TGG GGA GCA AAC ATC AAT TAT  
Gln Glu Asn Asn Tyr Glu Leu Ile Asn Leu Phe Thr Glu Trp Gly Ala Asn Ile Asn Tyr

VI  
301 AGC TTA CTC ACT CTC AAT ACG GAG CAT ACT TCG AAT TTC TGC CGG GAG TTA GGC GCA ---  
Ser Leu Val Thr Val Asn Thr Glu His Thr Trp Asn Phe Cys Arg Glu Leu Gly Ala

301 GCG TTC GTT ACT CTC AAT ATG GAG TGC ACC CAG CAT CTA TGC CAA AAG --- TTA GGT GCG  
Gly Leu Val Thr Val Asn Met Glu Cys Thr Gln Asp Leu Cys Gln Lys Leu Gly Ala

301 GGA TTA ATT TCC ATT AAT ACT GAG CAT GCC CGG GAC CTA TGT CGA AAA TTA GGA GCT ---  
Gly Leu Ile Ser Ile Asn Thr Glu His Ala Arg Asp Leu Cys Arg Lys Leu Gly Ala

361 AAA ATT TTC AAT CAA --- ATG GAT ATT GTA CAA ATA TTT TAT AAA ATA CAT CGT ATT AAA  
Lys Ile Leu Asn Glu Met Asp Ile Val Gln Ile Phe Tyr Lys Ile His Arg Ile Lys

361 ACG AAA GCT TTG AGT GAA AAT AAA ATT TTA GAA ATA TTT TAT AAT GTA CAG TAT GTT AAA  
Arg Lys Ala Leu Ser Glu Asn Lys Ile Leu Glu Ile Phe Tyr Asn Val Gln Tyr Val Lys

361 AAA GAA ATG CTT GAA AGA AAT GAA GTT ATA CAA ATT GTA TTT AAA ACA TTA GAT GAT ATC  
Lys Glu Met Leu Glu Arg Asn Glu Val Ile Gln Ile Val Phe Lys Thr Leu Asp Asp Ile

VII  
421 ACT AGT AGT AAT ATT ATT CTA TGC CAT AAA TTG TTA TCT AAT AAT CCC CTT TTC CAG AAT  
Thr Ser Ser Asn Ile Ile Leu Cys His Lys Leu Leu Ser Asn Asn Pro Leu Phe Gln Asn

421 ACT ACG AGT AAT ATT ATT CTA TGC CAT GAA TTA TTA TCT GAT AAC CCC CTA TTC CTA AAT  
Thr Ser Ser Asn Ile Ile Leu Cys His Glu Leu Leu Ser Asp Asn Pro Leu Phe Leu Asn

421 ACC AGT AGT AAT ATA ATT TTA TGT CAT GAA TTA TTT ACC AAC AAT CCT CTT TTA GAG AAT  
Thr Ser Ser Asn Ile Ile Leu Cys His Glu Leu Phe Thr Asn Asn Pro Leu Leu Glu Asn

VIII  
481 ATA GAG GAA TTA AAA ATA ATA ATT TGT TGT CTA GAG AAG ATA TCG ATC AAC TTT ATA  
Ile Glu Glu Leu Lys Ile Ile Cys Cys Phe Leu Glu Lys Ile Ser Ile Asn Phe Ile

481 AAT GCT CAA CTG AAA TTA AGA ATT TTT GGT GAA CTA GAT ACA TTA TCA ATC AAC TTT ACA  
Asn Ala Gln Leu Lys Leu Arg Ile Phe Gly Glu Leu Asp Thr Leu Ser Ile Asn Phe Thr

481 GTA AAT ATG GGG GAA ATG AGG ATG ATA ATT CAT TGG AGG ATG AAA AAT TTA ACG AAC CTA  
Val Asn Met Gly Glu Met Arg Met Ile Ile His Trp Arg Met Lys Asn Leu Thr Asn Leu

541 TTG GAT AAT ATT --- TCA TTC AAC GAA ATG CTA ACT AGG TAT TGG TAT AGT ATG GCG ATA  
Leu Asp Asn Ile Ser Phe Asn Glu Met Leu Thr Arg Tyr Trp Tyr Ser Met Ala Ile

541 TTA TTA AAT AAT GAC TCT ATT AGT GAA ATA TTA ACT AAA TTC TGG TAT GGT ATA GCA GTA  
Leu Leu Asn Asn Asp Ser Ile Ser Glu Ile Leu Thr Lys Phe Trp Tyr Gly Ile Ala Val

601 CGA TAT CAC CTA ACT GAA GCT ATC CAA TAT TTT TAT CAA CGT TAT AGA CAT TTT AAA GAT  
Arg Tyr His Leu Thr Ala Ile Gln Tyr Phe Tyr Gln Arg Tyr Arg His Phe Lys Asp

601 CTA TAT AAG CTT ACT GAA GCC ATC CAA TAT TTT TAT CAA CCA TAT AGT CAT TTT AAA GAT  
Leu Tyr Lys Leu Thr Glu Ala Ile Gln Tyr Phe Tyr Gln Pro Tyr Ser His Phe Lys Asp

601 AAA TAT AAT CTT AAG GAT GCA ATC CAA TAT TTT TAC CAG AGA TTC ATG CAC TTC AAC GAG  
Lys Tyr Asn Leu Lys Asp Ala Ile Gln Tyr Phe Tyr Gln Arg Phe Met Asp Phe Asn Glu

661 TGG CGG CTA ATA TGT GGG CTT TCT TTT AAT AAC GTG TCC GAT CTT CAT GAA ATA TAT CAT  
Trp Arg Leu Ile Cys Gly Leu Ser Phe Asn Asn Val Ser Asp Leu His Glu Ile Tyr His

661 TGG CGG TTA ATA TGT GGG GTT GCT TAT AAC AAC GTC TTT GAC CTT CAT GAA ATT TAT AAC  
Trp Arg Leu Ile Cys Gly Val Ala Tyr Asn Asn Val Phe Asp Leu His Glu Ile Tyr Asn

661 TGG CGA GTA ACA TGT GCT CTT TCT TTT AAT AAT GTG AAT GAT CTT CAT AAG ATG TAT ATA  
Trp Arg Val Thr Cys Ala Leu Ser Phe Asn Asn Val Asn Asp Leu His Lys Met Tyr Ile

721 ATA AAA AAG GTT GAT ATG AAT ATT GAT GAA ATG ATG TAT CTG GCC TGT ATG AGA --- GAT  
Ile Lys Lys Val Asp Met Asn Ile Asp Glu Met Met Tyr Leu Ala Cys Met Arg --- Asp

721 AAA CAG AAG ACG AAT ATA GAC ATT GAT GAA ATG ATG CAG TTG GCC TGT ATG TAT --- GAT  
Lys Glu Lys Thr Asn Ile Asp Ile Asp Glu Met Met Gln Leu Ala Cys Met Tyr --- Asp

721 ACA GAG AAG GTT CAT ACG AAT AAT GAC GAA ATG ATG AAT CTA GCC TGC AGC ATT CAA GAC  
Trp Arg Leu Thr Thr Ile Tyr Thr Cys Asn Asp Glu Met Met Asn Asp Leu His Ser Ile Tyr Asn

781 AGC AAT TTT TTA ACC ATT TTC TAT TGT TTT GTA TTA TTA GGT AAC ATC AAT CCG GCA ATG  
Ser Asn Phe Leu Thr Ile Phe Tyr Cys Phe Val Leu Gly Ala Asn Ile Asn Arg Ala Met

781 TGT AAT TAT ACA ACT ATA TAT TAT TGT TGT ATG TTG GGA GCT GAC ATT AAT CCG GCA ATG  
Cys Asn Tyr Thr Thr Ile Tyr Tyr Cys Cys Met Leu Gly Ala Asp Ile Asn Arg Ala Met

781 AGA AAT TTA TCA ACC ATT TAC TAT TGT TTT CTA TTG GGG GCT AAC ATC AAT CAA GCA ATG  
Arg Asn Leu Ser Thr Ile Tyr Tyr Cys Phe Leu Leu Gly Ala Asn Ile Asn Gln Ala Met

841 GTT ACT TCG GTA AAA AAG TTT TAT ACT AAT AAC TTG TTC TTT TGT ATA GAT TTA GGA GCT  
Val Thr Ser Val Lys Asn Phe Tyr Thr Asn Asn Leu Phe Phe Cys Ile Asp Leu Gly Ala

841 ATT ACC TCG GTA AAT ATT TTT TGT GAA GGT AAC TTA TTT TGT GAT TTA GGA GCT  
Ile Thr Ser Val Met Asn Phe Cys Glu Gly Asn Leu Phe Leu Cys Met Asp Leu Gly Ala

841 TTG ACC TCA GTA TTA AAT TAT AAT ATT TTT AAC TTA TTC TTT TGT ATA GAC TTA GGG GCT  
Leu Thr Ser Val Leu Asn Tyr Asn Ile Phe Asn Leu Phe Phe Cys Ile Asp Leu Gly Ala

901 AAT GCT TTC CAA GAG AGC TTA GAA TTA GCA AAA CAA AAG AAT CAT GAT ATA TTA GTA GAA  
Asn Ala Phe Glu Glu Ser Leu Glu Leu Ala Lys Gln Lys Asn His Asp Ile Leu Val Glu

901 GAT GCG TTT GAA GAG AGC ATG GAA ATA GCG AGT CAA ACG AAT AAT TGG ATA TTA ATA AAT  
Asp Ala Phe Glu Glu Ser Met Glu Ile Ala Ser Gln Thr Asn Ser Thr Phe Ser Ser

901 GAT GCC TTT GAA GAG GGT AAG ACC CTC GCG AAA CAA AAG GGG TAT AAT GAA ATA CTG GAA  
Asp Ala Phe Glu Glu Gly Lys Thr Leu Ala Lys Gln Lys Gly Tyr Asn Glu Ile Val Glu

961 ATA TTA TCA TTT AAA GAT TTT TAT AAT TCA AAC GTC TCT CTT TTA TCT TTA AAA ACG ACA  
Ile Leu Ser Phe Lys Asp Phe Tyr Asn Ser Asn Val Ser Leu Leu Ser Leu Lys Thr Thr

961 ATC TTA TTA TTT AAA AAT TAC AGT CCA GAC TCT TCT CTT TTA TCA ATA AAA ACG --- ACA  
Ile Leu Ser Leu Phe Lys Asn Tyr Ser Pro Asp Ser Ser Leu Leu Ser Ile Lys Thr --- Thr

961 ATC TTA TCA TTA GAT ATC ATT TAT AGT CCA AAT ACT GAC TTC TCA TCA --- AAA ATA ---  
Ile Leu Ser Leu Asp Ile Ile Tyr Ser Pro Asn Thr Asp Phe Ser Ser Ser Ser Ser

1021 GAT CCA GAA AAA ATT AAT GCC TTG TTA AAA --- AAC --- TAT AGA TCT AAA AAT ATA ATG  
Asp Pro Glu Lys Ile Asn Ala Leu Leu Lys Tyr Arg Ser Lys Asn Ile Met

1021 GAT CCG GAA AAA ATT AAC GCC TTA CTA GAT GAA GAA AAG TAT AAG TCA AAA AAT ATG TTA  
Asp Pro Glu Lys Ile Asn Ala Leu Leu Asp Glu Glu Lys Tyr Lys Ser Lys Asn Met Leu

1021 GAA CCT GAA CAT ATT AGT TCT TTG TTA AAA --- AAC TTT TAT --- CCA AAA AAT CTG TTC  
Glu Pro Glu His Ile Ser Ser Leu Leu Lys Asn Phe Tyr Pro Lys Asn Leu Phe

1081 AGG TAT AAA AAG TTG TGT CCG AAA ATA ATA CCG TCG GCG ACA TTT ATT ATA  
Arg Tyr Lys Lys Leu Cys Pro Lys Ile Ile Arg Trp Ala Arg Phe Ile Ile

1081 ATA TAT GAA GAA TCA TTG TTT CAC ATC TAT --- GGG GTA AAC --- ATT  
Ile Tyr Glu Glu Ser Leu Phe His Ile Tyr Gly Val Asn Ile

1081 GCT TTT GAT CGT --- TGC --- AAT CCT --- --- GGT TTA TAT TAT TCT  
Ala Phe Asp Arg Cys Asn Pro Gly Leu Tyr Tyr Ser

Figura 25. Comparación de las secuencias de nucleótidos de las fases de lectura abierta D'42, K'362, K'360 y D'363 y su traducción a aminoácidos. Los guiones indican interrupciones introducidas para maximizar la homología. El símbolo indica identidad de aminoácido en posiciones análogas de las secuencias de K'362, K'360 y D'363. El símbolo o indica aminoácidos homólogos según los grupos: Ser, Thr, Pro, Ala, Gly; Met, Ile, Leu, Val; Asp, Glu, Asn, Gln; Arg, Lys, His; Phe, Tyr, Trp (172). Los límites entre segmentos con distinto grado de homología ( en números romanos) se denotan con trazos verticales.



A

NUCLEOTIDOS			AMINOACIDOS		
	K'360	K'362		K'360	K'362
K'362	273		K'362	143	
D'363	399	384	D'363	242	186

	D'42	K'362	K'360		D'42	K'362	K'360
K'362	11			K'362	4		
K'360	26	24		K'360	13	12	
D'363	33	35	30	D'363	20	18	18

B

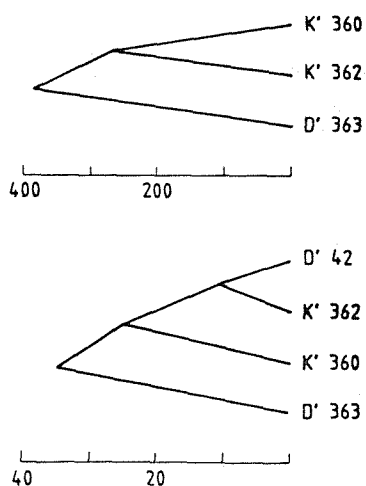


Figura 26. Cuantificación de la divergencia de secuencia existente entre las fases de lectura abierta D'42, K'360, K'362 y D'363. A) Matrices de diferencias en la secuencia de nucleótidos o aminoácidos. Se cuantificaron siguiendo los criterios de Fitch y Margoliash (131). La parte superior corresponde a la comparación de las secuencias completas de K'360, K'362 y D'363 y la inferior a los 30 primeros codones de las cuatro fases. B) Dendrogramas calculados a partir de los datos de divergencia de secuencia de nucleótidos de las tres (parte superior) o de las cuatro fases (parte inferior).

TABLA VII. Frecuencias de uso de codones de las proteínas correspondientes a la traducción de las fases de lectura abierta K'362, K'360 y D'363.

	T			C			A			G			
T	Phe	17	13	12	Ser	9	4	5	Tyr	15	18	16	T
		5	4	7		2	0	2		1	1	2	C
	Leu	18	22	24		3	5	5		Trp	7	6	A
		11	7	5		2	1	0			6	6	G
C	Leu	6	5	8	Pro	0	1	3	His	8	8	10	T
		1	0	0		1	1	0		3	1	0	C
		6	9	5		3	4	4	Gln	5	7	7	A
		3	2	2		1	1	1		4	5	2	G
A	Ile	13	14	12	Thr	8	8	7	Asn	22	18	28	T
		4	5	7		2	3	5		11	10	9	C
		22	14	15		3	7	3	Lys	23	16	13	A
	Met	13	14	4		2	4	2		4	6	8	G
G	Val	2	4	2	Ala	9	7	8	Asp	13	13	8	T
		3	2	0		3	4	4		1	6	7	C
		5	5	6		6	5	5	Glu	13	19	20	A
		2	1	4		2	4	1		8	5	8	G

La primera, segunda y tercera base de cada codón corresponden a las representadas en los márgenes izquierdo, superior y derecho, respectivamente. Las tres cifras junto a cada codon corresponden, de izquierda a derecha, a las fases K'362, K'360 y D'363.

TABLA VIII. Segmentos con distinto grado de homología de secuencia en las fases K'362, K'360 y D'363.

SEGMENTO	K'362-K'360		K'362-D'363		K'360-D'363	
	N	A	N	A	N	A
I	64	69	66	62	66	86
II	37	40	35	23	35	12
III	83	79	83	86	81	86
IV	67	43	67	32	51	26
V	89	89	82	85	82	81
VI	53	46	58	37	51	34
VII	88	85	67	75	72	75
VIII	72	48	44	9	38	21
IX	78	32	73	59	67	53
X	87	80	87	93	89	87
XI	61	53	49	33	46	22

Se muestra la homología porcentual que presentan distintos segmentos (figura 25) de las secuencias de nucleótidos (N) y aminoácidos (A), comparados por parejas. Se cuantificó el número de diferencias (residuos diferentes, o ausentes por la introducción de huecos para maximizar la homología) en ambos tipos de secuencias y se dividió por la longitud total del segmento (incluyendo los huecos).

mientras que el segmento II presenta el valor más bajo. Las zonas restantes tienen un grado de homología intermedio, diferente si comparamos nucleótidos o aminoácidos y dependiendo del par de secuencias que se comparen.

Los perfiles de hidropatía de las proteínas correspondientes a las fases K'360, K'362 y D'363 se muestran comparados en la figura 27. En esta figura se muestra asimismo la posición de los segmentos con distinto grado de homología de secuencia. Los segmentos con mayor homología contienen zonas predichas como internas en las estructuras tridimensionales de las proteínas. Todas las zonas predichas como expuestas al solvente se encuentran en segmentos con menor homología de secuencia.

Los extremos N-terminales de las tres secuencias tienen un carácter moderadamente hidrofóbico. Sin embargo, no resultan predichas como péptidos señal utilizando el método de McGeoch (141).

Las estructuras secundarias predichas a partir de las secuencias de aminoácidos de K'360, K'362 y D'363 se representan en la figura 28. La estructura general es similar en los tres casos. Presentan contenidos globales elevados de hélice  $\alpha$  y estructura  $\beta$  (alrededor de un 40% y un 30%, respectivamente). Las estructuras  $\alpha$  y  $\beta$  predichas se distribuyen a lo largo de toda la molécula en los tres casos. En esta misma figura se representan las posiciones de los residuos de cisteína y de los sitios potenciales de N-glicosilación.

Con el fin de predecir la localización de epitopos continuos en las secuencias de K'360, K'362 y D'363 se evaluaron su hidrofiliidad y su antigenicidad potencial. El resultado de este análisis se muestra en la figura 29. Existen en cada secuencia varias zonas en las que los valores de ambos parámetros son significativamente altos. Estas zonas coinciden generalmente con segmentos con menor homología de secuencia, según se observa en la figura 29.

En la tabla IX se muestran algunas características de los sitios potenciales de N-glicosilación presentes en las secuencias de las tres fases. No existe ningún sitio localizado en posiciones análogas de las tres secuencias. Los valores de índice hidropático de los segmentos que incluyen los sitios no predicen que sus residuos se encuentren expuestos al solvente. Tampoco resultan predichos como incluidos en epitopos continuos a partir de los perfiles de hidrofiliidad y antigenicidad potencial. Todos estos datos y la estructura secundaria predicha en dichas zonas se recogen en la tabla IX.

El análisis de las secuencias individuales de aminoácidos revela algunos segmentos que presentan homología. Están centrados en residuos de cisteína. Sus posiciones y secuencias se muestran en la figura 30.

Se llevó a cabo una búsqueda de proteínas cuyas secuencias presentaran homología con las de la familia 360 con la ayuda del programa FASTP. En la colección de secuencias PROTSEQ no se encontró ninguna con un grado de homología sig-

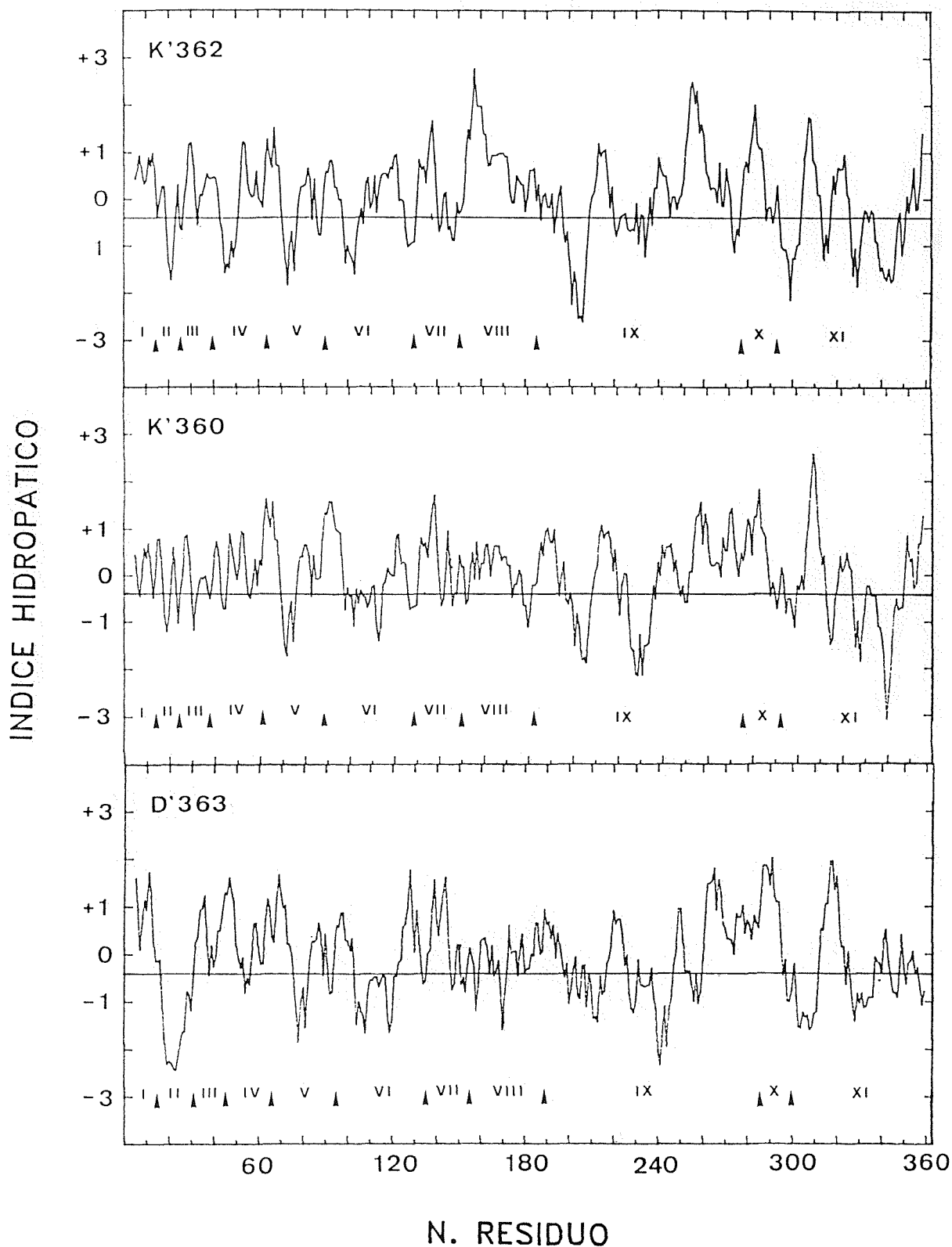


Figura 27. Carácter hidrofóbico de los segmentos con mayor homología de secuencia de las fases K'362, K'360 y D'363.

Se muestran los perfiles de hidropatía deducidos a partir de la secuencia de aminoácidos de las tres fases. Los valores positivos de índice hidropático corresponden a zonas predichas como internas en la estructura tridimensional de la proteína, y los negativos a las predichas como expuestas al solvente. Se señalan con flechas verticales en cada fase los límites entre los segmentos que presentan distinto grado de homología (figura 25), señalándose estos con números romanos.

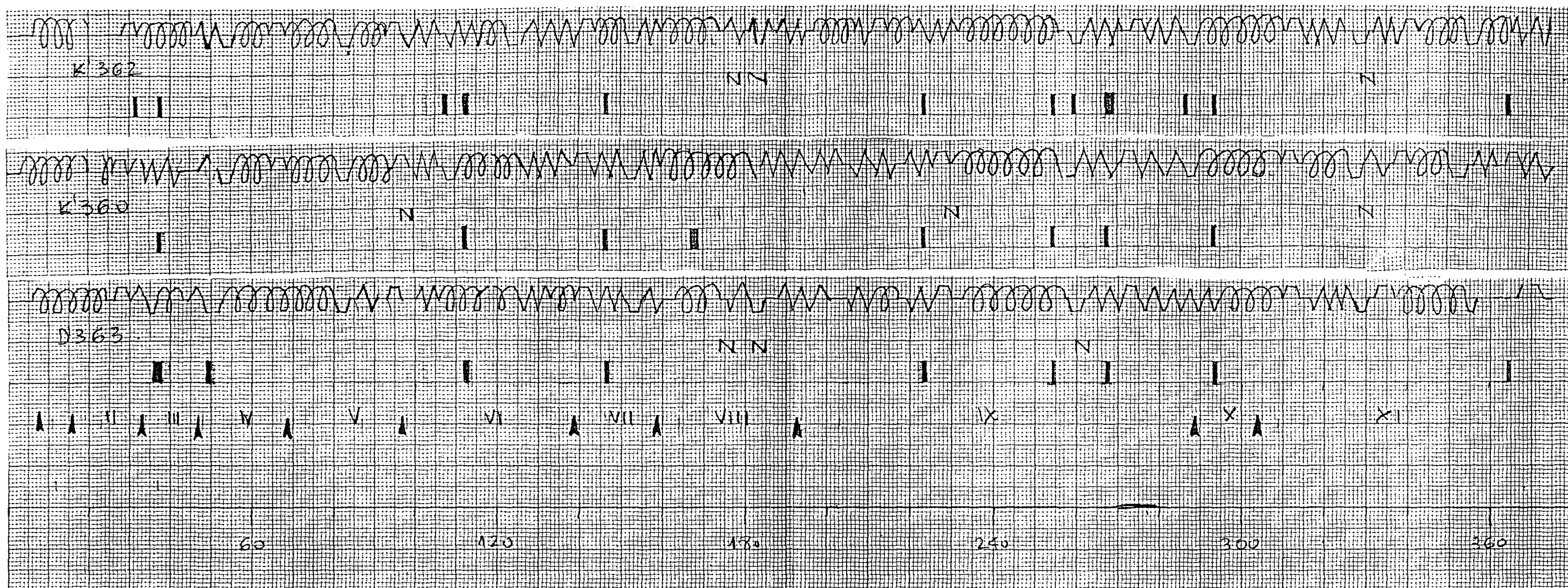


Figura 28. Predicción de estructura secundaria de las proteínas correspondientes a las fases K'362, '360 y D'363.

Los distintos tipos de trazos corresponden a las diferentes estructuras secundarias.  $\alpha$ , hélice  $\alpha$ ;  $\beta$ , cadena  $\beta$ ;  $\gamma$ , giro  $\beta$ ; ---, estructura aperiódica. Las interrupciones en el trazo reflejan los huecos introducidos para maximizar la homología entre las secuencias. La posición de los segmentos con distinto grado de homología se señalan con flechas verticales. Los símbolos  $\square$  y N representan los residuos de cisteína y los sitios potenciales de N-glicosilación, respectivamente. En la parte inferior se muestran los porcentajes predichos de cada tipo de estructura secundaria:  $\alpha$ , hélice  $\alpha$ ;  $\beta$ , cadena  $\beta$ ;  $\gamma$ , giro  $\beta$ ; C, estructura aperiódica.



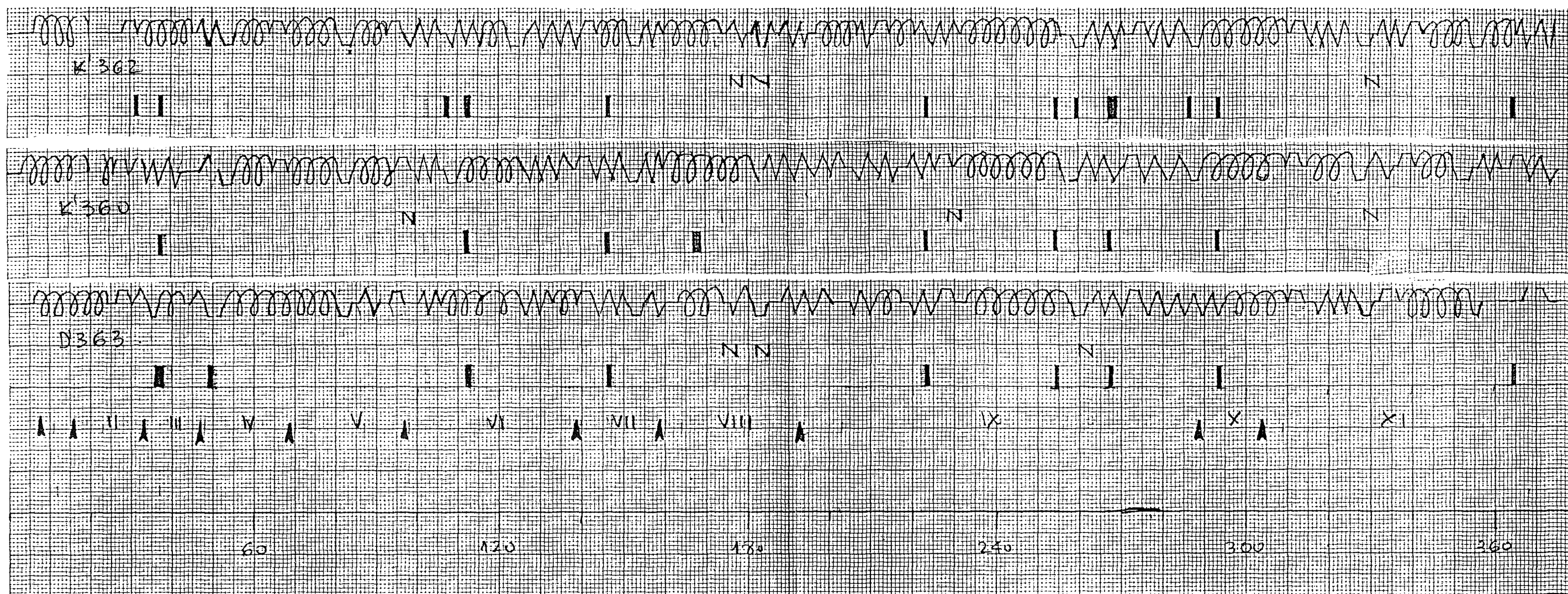


Figura 28. Predicción de estructura secundaria de las proteínas correspondientes a las fases K'362, '360 y D'363.

Los distintos tipos de trazos corresponden a las diferentes estructuras secundarias.  $\alpha$ , hélice  $\alpha$ ;  $\beta$ , cadena  $\beta$ ;  $\gamma$ , giro  $\beta$ ; ---, estructura aperiódica. Las interrupciones en el trazo reflejan los huecos introducidos para maximizar la homología entre las secuencias. La posición de los segmentos con distinto grado de homología se señalan con flechas verticales. Los símbolos  $\square$  y N representan los residuos de cisteína y los sitios potenciales de N-glicosilación, respectivamente. En la parte inferior se muestran los porcentajes predichos de cada tipo de estructura secundaria:  $\alpha$ , hélice  $\alpha$ ;  $\beta$ , cadena  $\beta$ ; T, giro  $\beta$ ; C, estructura aperiódica.



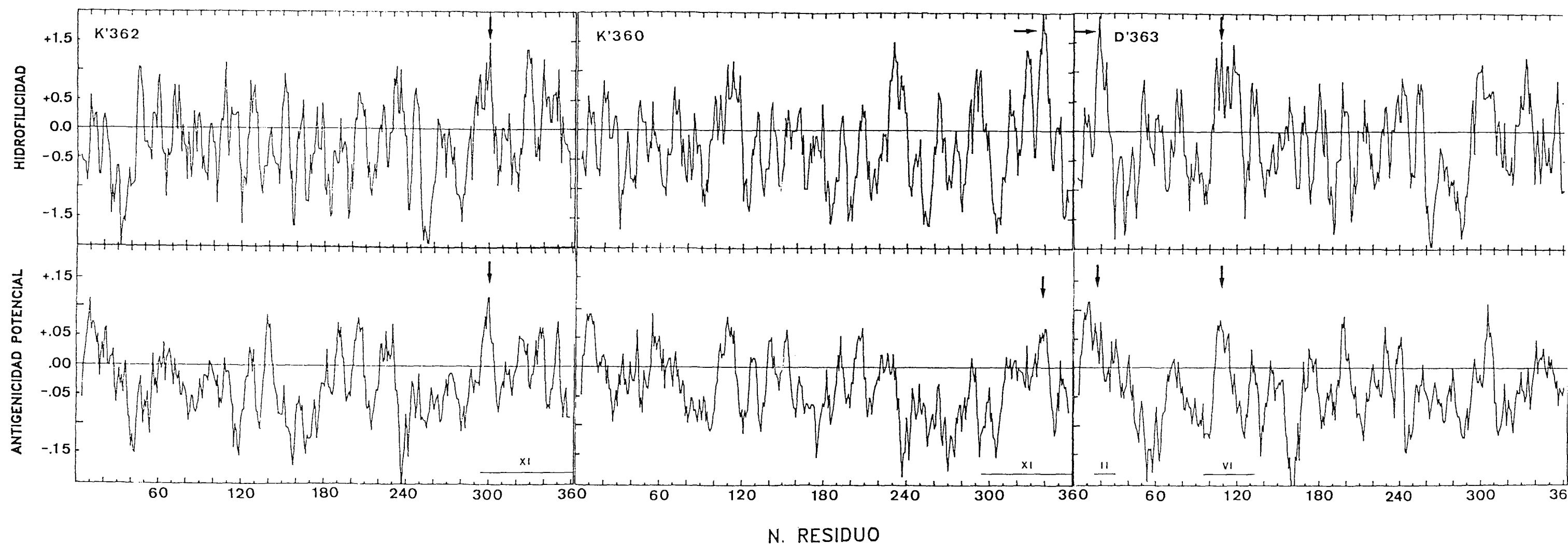


Figura 29. Perfiles de hidrofiliidad y antigenicidad potencial de las proteínas correspondientes a las fases de lectura abierta K'362, K'360 y D'363.

Se muestran comparados los perfiles para cada fase. En ordenadas, los parámetros hidrofiliidad y antigenicidad potencial asignados, respectivamente, al cuarto residuo de cada segmento de seis y al cuarto de cada segmento de siete. En abscisas, el número de residuo. Los valores más altos de hidrofiliidad y de antigenicidad potencial superiores a 0.05 corresponden a segmentos predichos como participando en epitopos continuos. Las flechas señalan la posición del valor más alto de hidrofiliidad (o de los dos valores más altos en el caso de D'363). Las zonas señaladas presentan valores de antigenicidad potencial superiores a 0.05. Estas zonas se encuentran localizadas en segmentos con menor homología de secuencia. La posición de cada segmento (denominado con números romanos) se muestra con trazos horizontales.

TABLA IX. Sitios potenciales de N-glicosilación en las fases K'362, K'360 y D'363.

K'362	90	219	318
HID	+/-	+/-	-
ANT	-	+/-	-
HIP	-	-	-
SEC	T	T	T
K'360	167	172	313
HID	-	-	+/-
ANT	-	-	-
HIP	-	-	-
SEC	$\beta$	$\beta$	T
D'363	171	179	257
HID	+/-	+/-	+/-
ANT	-	+/-	+/-
HIP	+/-	-	+/-
SEC	$\beta$	T	T

Se muestran la posición de su primer residuo y sus valores de hidrofiliidad (HID), antigenicidad potencial (ANT) e índice hidropático (HIP). El símbolo - indica valores negativos de hidrofiliidad o antigenicidad potencial y positivos de índice hidropático en los residuos correspondientes al sitio potencial. El símbolo +/- indica valores positivos de hidrofiliidad y antigenicidad potencial o negativos de índice hidropático en los dos residuos flanqueantes al sitio. Se indica asimismo la estructura secundaria predicha:  $\beta$ , cadena  $\beta$ ; T, giro  $\beta$ .

105

T	W	N	F	C	R	E	L	G	A	- K I	K'362
T	Q	D	L	C	Q	K	L	G	A	R K A	K'360
A	R	D	L	C	R	K	L	G	A	- K E	D'363

137

N	I	I	L	C	H	K	L	L	S	N	- N	K'362
N	I	I	L	C	H	E	L	L	S	E	- N	K'360
N	I	I	L	C	H	E	L	F	T	N	- N	D'363

256

T	I	F	Y	C	F	V	L	G	A	N	I	N	K'362
T	I	Y	Y	C	C	M	L	G	A	D	I	N	K'360
T	I	Y	Y	C	F	L	L	G	A	N	I	N	D'363

282

N	L	-	F	C	I	D	L	G	A	N	A F	K'362
N	L	-	L	C	M	D	L	G	A	D	A F	K'360
N	L	-	L	C	I	D	L	G	A	D	A F	D'363

Figura 30. Homología interna en las secuencias de aminoácidos de las fases '362, K'360 y D'363.

Se muestran las secuencias de las zonas homólogas en código de una letra, correspondiendo la numeración a la de K'362. Los guiones corresponden a huecos introducidos para maximizar la homología entre las distintas zonas. En cada posición análoga se ha contabilizado el número de residuos homólogos de acuerdo con los grupos utilizados en la figura 25. En cada posición se han rodeado con trazos los residuos homólogos presentes en cinco o más zonas homólogas.

nificativo de acuerdo con los criterios establecidos por Lipman y Pearson(138).

### 5.2. K'L177.

La secuencia de nucleótidos de esta fase de lectura abierta y su traducción a aminoácidos se muestran en la figura 31. La composición de aminoácidos, el peso molecular y el punto isoelectrico teórico de la proteína que se sintetizaría si la fase se tradujese desde el primer codon de iniciación se muestran en la tabla X. La frecuencia de uso de codones correspondiente se resume en al tabla XI.

La predicción de exposición al solvente de las distintas zonas de la secuencia por medio del índice hidropático se muestra en la figura 32. La zona centrada en el residuo 20, por su carácter extremadamente hidrofóbico y su longitud podría tener una funcion de péptido señal. Esta secuencia se señala en la figura 31. Las secuencias Gly-Ser-Gly y Val-Arg-Gly situadas más allá de la secuencia hidrofóbica en dirección C-terminal se ajustan a la secuencia consenso reconocida por endopeptidasas. Su posición se señala en la figura 31.

En la figura 33 se representa la estructura secundaria predicha a partir de su secuencia de aminoácidos. Los contenidos globales de estructuras  $\alpha$  y  $\beta$  no son elevados: 26 y 16%, respectivamente. Abundan los giros  $\beta$ , así como las zo-

nas predichas en estructura aperiódica (29% en ambos casos). Posee un elevado número de residuos de cisteína (10), agrupados en dos zonas. No posee sitios potenciales de N-glicosilación.

Los perfiles de hidrofiliicidad y antigenicidad potencial deducidos a partir de su secuencia de aminoácidos se representan en la figura 34. Existen al menos dos zonas en las que los valores de ambos parámetros predicen de forma significativa la presencia de epitopos continuos.

									30										60
Met	Phe	Asn	Ile	Lys	Met	Thr	Ile	Ser	Thr	Leu	Leu	Ile	Ala	Leu	Ile	Ile	Leu	Val	Ile
ATG	TTT	AAT	ATT	AAA	ATG	ACA	ATT	TCT	ACA	TTG	CTT	ATT	GCT	CTT	ATT	ATA	CTA	GTT	ATT
					→				90										120
Ile	Ile	Leu	Val	Val	Phe	Leu	Tyr	Tyr	Lys	Lys	Gln	Gln	Pro	Pro	Lys	Lys	Val	Cys	Lys
ATT	ATT	TTA	GTC	GTA	TTT	TTA	TAT	TAT	AAA	AAA	CAA	CAA	CCA	CCA	AAA	AAG	GTC	TGT	AAA
								←	150										180
Val	Asp	Lys	Asp	Cys	Gly	Ser	Gly	Glu	His	Cys	Val	Arg	Gly	Thr	Cys	Ser	Thr	Leu	Ser
CTA	GAT	AAA	GAT	TGT	GGT	AGT	GGA	GAG	CAT	TGT	GTT	CGT	GGA	ACA	TGC	AGC	ACA	TTG	AGC
									210										240
Cys	Leu	Asp	Ala	Val	Lys	Met	Asp	Lys	Arg	Asn	Ile	Lys	Ile	Asp	Ser	Lys	Ile	Ser	Ser
TGC	TTA	GAC	GCT	GTA	AAA	ATG	GAC	AAA	CGA	AAT	ATT	AAG	ATA	GAT	TCT	AAG	ATT	TCC	TCA
									270										300
Cys	Glu	Phe	Thr	Pro	Asn	Phe	Tyr	Arg	Phe	Thr	Asp	Thr	Ala	Ala	Asp	Glu	Gln	Gln	Glu
TGT	GAA	TTC	ACT	CCC	AAT	TTT	TAC	CGT	TTT	ACG	GAT	ACT	GCT	GCC	GAT	GAG	CAG	CAA	GAA
									330										360
Phe	Gly	Lys	Thr	Arg	His	Pro	Ile	Lys	Ile	Thr	Pro	Ser	Pro	Ser	Glu	Ser	His	Ser	Pro
TTT	GGA	AAA	ACA	CGG	CAT	CCT	ATA	AAA	ATA	ACT	CCA	TCT	CCA	AGT	GAA	TCC	CAT	AGC	CCC
									390										420
Gln	Glu	Val	Cys	Glu	Lys	Tyr	Cys	Ser	Trp	Gly	Thr	Asp	Asp	Cys	Thr	Gly	Trp	Glu	Tyr
CAA	GAG	GTG	TGT	GAA	AAA	TAT	TGT	TCA	TGG	GGA	ACC	GAT	GAC	TGT	ACA	GGT	TGG	GAA	TAT
									450										480
Val	Gly	Asp	Glu	Lys	Glu	Gly	Thr	Cys	Tyr	Val	Tyr	Asn	Asn	Pro	His	His	Pro	Val	Leu
GTT	GGT	GAT	GAA	AAG	GAG	GGA	ACA	TGT	TAT	GTA	TAT	AAT	AAT	CCA	CAT	CAC	CCG	GTT	CTT
									510										
Lys	Tyr	Gly	Lys	Asp	His	Ile	Ile	Ala	Leu	Pro	Arg	Asn	His	Lys	His	Ala			
AAA	TAT	GGT	AAG	GAT	CAC	ATC	ATA	GCC	TTA	CCT	AGA	AAT	CAT	AAA	CAT	GCA			

Figura 31. Secuencias de nucleótidos y aminoácidos de la fase K'L177. Se señala con flechas horizontales la posición del segmento hidrofóbico. Las secuencias subrayadas se ajustan a la consenso de reconocimiento por endopeptidasas. La señalada con una flecha vertical es una secuencia realmente reconocida en una proteína (142).

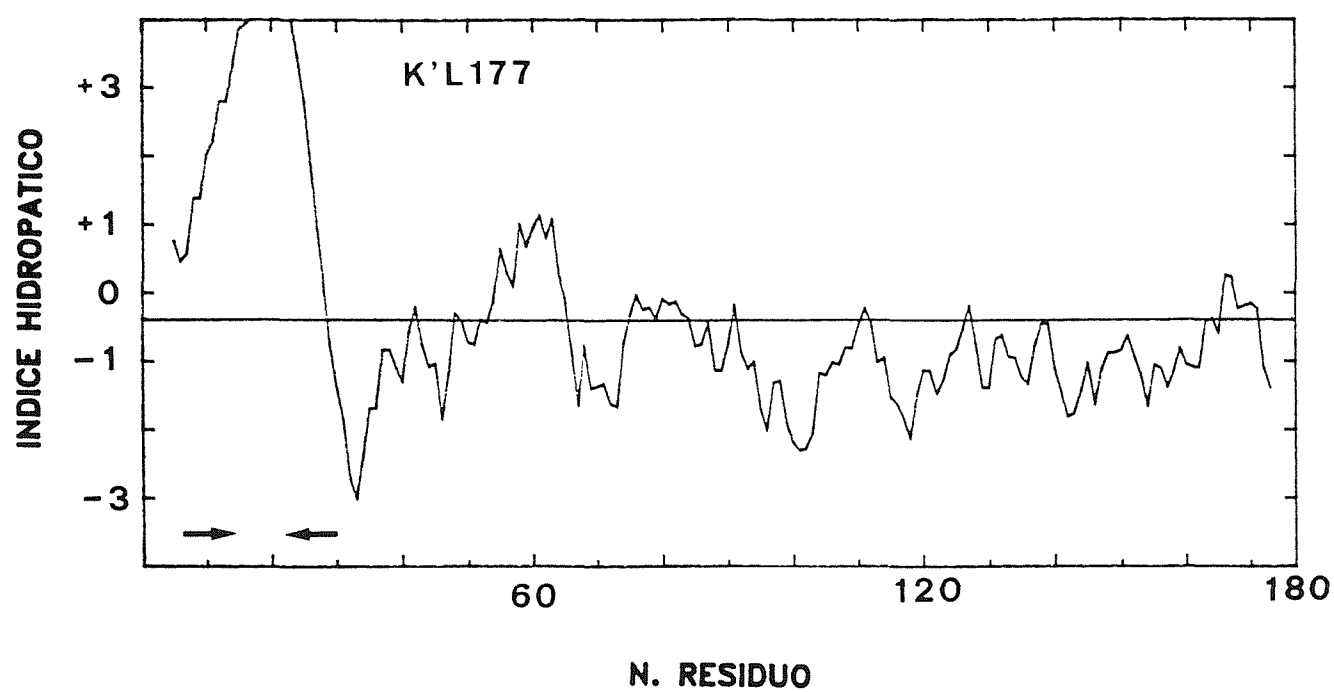


Figura 32. Perfil de hidropatía de la fase K'L177.

Se señala con flechas horizontales el segmento de carácter fuertemente hidrofóbico que resulta predicho como péptido señal.

TABLA X. Composición de aminoácidos de la proteína corres-  
pondiente a la fase K'L177.

A	C	D	E	F	G	H	I	K	L
6	10	11	10	6	9	8	15	18	10

M	N	P	Q	R	S	T	V	W	Y
3	6	10	5	5	12	12	11	2	8

Pm            20.160

pI            7.75

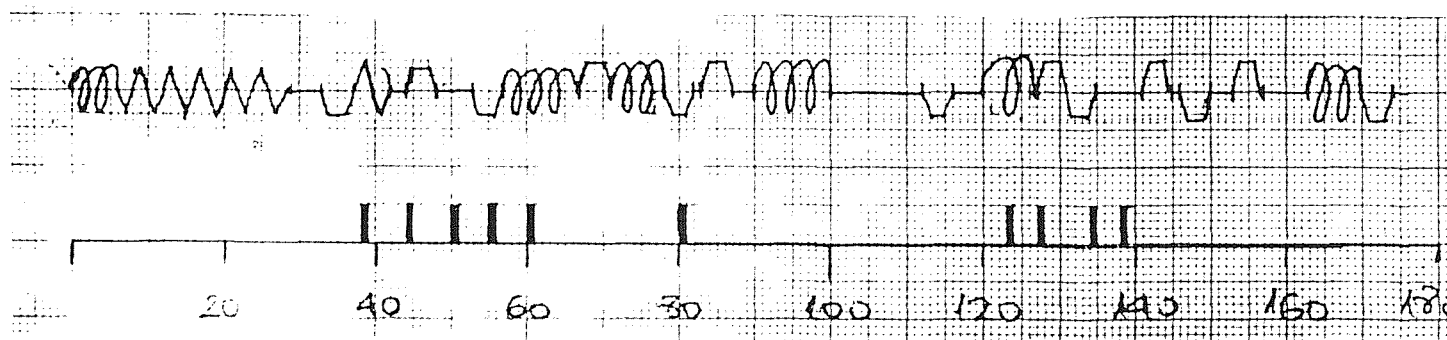
Se muestran asimismo su peso molecular y su punto isoeléctrico teórico.



TABLA XI. Frecuencia de uso de codones de la proteína co-  
rrespondiente a la fase K'L177.

	T		C		A		G		
T	Phe	5	Ser	3	Tyr	7	Cys	8	T
		1		2		1		2	C
	Leu	4		2					A
		2		0			Trp	2	G
C	Leu	3	Pro	2	His	6	Arg	2	T
		0		2		2		0	C
		1		5	Gln	4		1	A
		0		1		1		1	G
A	Ile	9	Thr	3	Asn	6	Ser	2	T
		1		1		0		3	C
		5		7	Lys	13	Arg	1	A
	Met	3		1		5		0	G
G	Val	4	Ala	3	Asp	8	Gly	4	T
		2		2		3		0	C
		4		1	Glu	6		5	A
		1		0		4		0	G

La disposición de las bases de cada codon, como en la ta-  
bla VII.



%	$\alpha$	$\beta$	T	C
	26	16	19	29

Figura 33. Predicción de estructura secundaria de la proteína correspondiente a la fase K'L177.

Los tipos de estructura secundaria y la posición de las citeínas se simbolizan como en la figura 27. Se muestra asimismo la proporción de cada tipo de estructura.

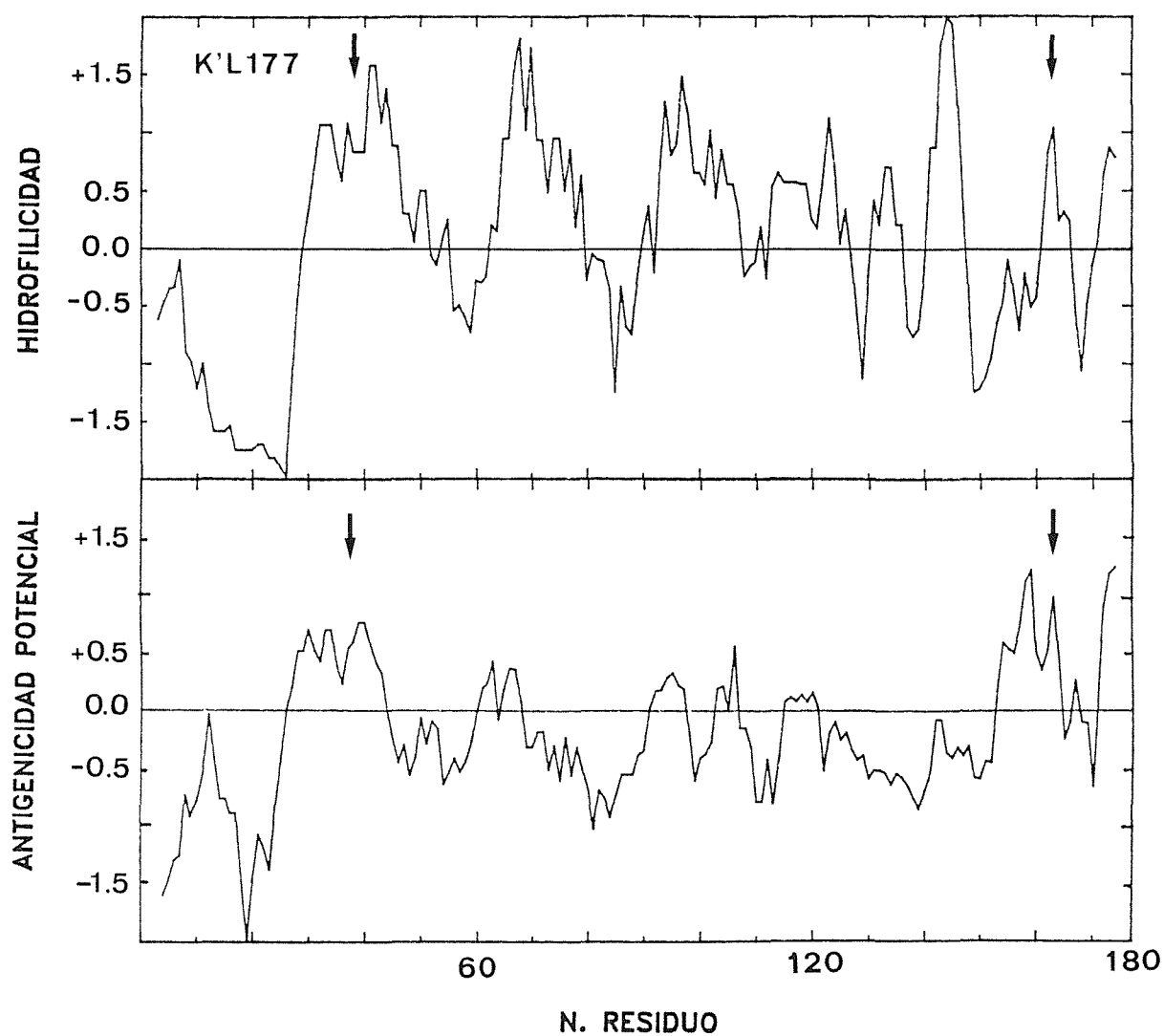


Figura 34. Perfiles de hidrofiliidad y antigenicidad potencial de la proteína correspondiente a la fase K'L177.

Se señalan con flechas verticales las zonas con valores más altos de hidrofiliidad y valores de hidrofiliidad superiores a 0.05.

## DISCUSSION

En este trabajo se describe la determinación de la secuencia de los fragmentos terminales ( EcoRI-K' y EcoRI-D' ) del genoma del VPPA, incluyendo la unión covalente terminal entre ambas cadenas del DNA, así como el análisis de la secuencia de nucleótidos obtenida.

### 1. HORQUILLA TERMINAL.

La secuencia de la unión covalente terminal se ha obtenido aprovechando la existencia de un sitio de restricción (BssHII) muy próximo al extremo del genoma. Suponiendo que la unión consistía en una horquilla terminal, como es el caso en el virus de la vacuna (54), este sitio permitiría la secuenciación de la horquilla utilizando el método de Maxam y Gilbert (72).

Un experimento de carácter analítico demostró la existencia en cada extremo del genoma de dos fragmentos BssHII de movilidad muy similar y correspondiente aproximadamente a la predicha para los fragmentos terminales (figura 10). La purificación y secuenciación de estos fragmentos (figuras 11 y 12) permitió concluir que la unión covalente terminal se puede describir como una horquilla, con lo que la molécula de DNA bicatenario consiste en una única cadena polinucleotídica ininterrumpida. Además, la unión existe en dos formas, cuyas secuencias son invertidas y complementarias (figura 15). Todas estas características de la unión covalente

terminal las exhibe asimismo la horquilla terminal del virus de la vacuna (54). Otras características comunes son su estructura secundaria conteniendo bases desapareadas (figura 15) y su alto contenido en A+T ( $>90\%$ ).

Sin embargo, estas dos horquillas terminales se diferencian por su tamaño, secuencia y grado de apareamiento de bases en su estructura secundaria más estable. La presente en el genoma del VPPA es de menor longitud (37 vs 104 nucleótidos) y contiene más bases desapareadas (35% vs 15%) que su homóloga del virus de la vacuna.

No existe homología de secuencia significativa entre ambas estructuras, pero sí en la secuencia adyacente a las mismas, según se muestra en la figura 35. Se trata de una zona de 16 nucleótidos de longitud que presenta una homología de un 80%. En la región adyacente a la horquilla existe asimismo un segmento cuya secuencia complementaria presenta homología con la secuencia del virus de la vacuna cuando ambas se escriben con la misma polaridad (figura 35). Es de reseñar finalmente la existencia de una secuencia palindrómica de 20 nucleótidos de longitud.

Estructuras similares a la presente en los extremos del genoma del VPPA y los ortopoxvirus se ha descrito para las formas replicativas del DNA monocatenario de los parvovirus (144,145) y del DNA mitocondrial bicatenario de *Paramecium* (146). Existe asimismo evidencia experimental de su presencia en el DNA ribosomal del ciliado *Tetrahimena* (147) y en el DNA cromosómico de levadura (148).

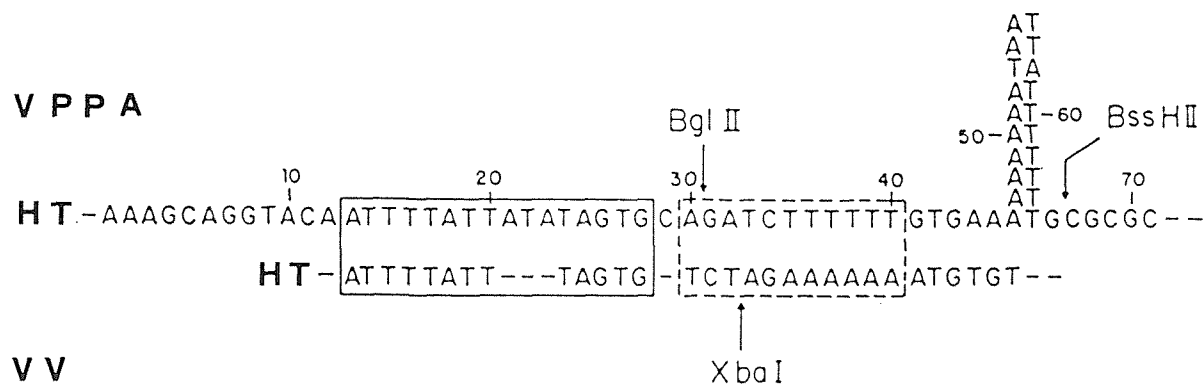


Figura 35. Homología de secuencia entre los genomas del VPPA y del VV en la zona adyacente a la horquilla terminal.

Rodeada con trazo continuo, una zona de homología de 16 nucleótidos. Rodeada con trazo discontinuo, una zona cuya secuencia complementaria con igual polaridad es idéntica. Se muestra, apareada, una zona palindrómica (46-65 en la numeración) así como algunos sitios de restricción (54). HT, horquilla terminal.

Las secuencias de los extremos de los DNAs de las parvovirus se han determinado, tanto en el caso de parvovirus autónomos como defectivos (149,150), y consisten en estructuras palindrómicas, que pueden adquirir una configuración estable en forma de horquilla, siendo la más estable en forma de Y en el caso del extremo 3' de los parvovirus autónomos (150,151), debido a su alto contenido en G+C. La longitud de estas horquillas es de aproximadamente 120 nucleótidos, bastante similar a los 104 nucleótidos que constituyen la horquilla terminal del virus de la vacuna.

Existen asimismo datos de secuencia sobre la zona no palindrómica del intermedio dimérico del DNA mitocondial de *Paramecium*. Esta zona corresponde a la horquilla terminal del monómero y sus rasgos más característicos son el alto contenido en A+T., la presencia de repeticiones directas en tandem y la reducida proporción de bases apareadas (152).

En cuanto al DNA ribosomal de *Tetrahimena*, solo existen datos de secuencia de la zona adyacente a la horquilla terminal, la cual se caracteriza por la presencia de un número variable de repeticiones en tandem de pequeño tamaño  $((C_4A_2)_n, n=20-70)$  y de la existencia de interrupciones específicas en, ambas cadenas (147).

La existencia de estas interrupciones ha permitido identificar secuencias repetidas teloméricas similares en tripanosomas y protozoos flagelados relacionados (153). Todas las secuencias teloméricas repetitivas encontradas en eucariotas inferiores se ajustan a la fórmula general



5'(C<sub>1-8</sub>(A/T<sub>1-4</sub>))3'(154).

El hecho de que estas características parecen ser comunes a los telómeros de los cromosomas de levadura (155,156) sugiere que los rasgos estructurales requeridos para la replicación y la resolución de telómeros se han conservado muy fielmente durante la evolución, dada la distancia evolutiva que media entre ambos organismos.

Aunque las dos formas de horquilla terminal (lenta y rápida, l y r) se encuentran presentes en cada extremo del genoma del VPPA en cantidades aproximadamente equimoleculares si analizamos una población de moléculas (figura 10), no se conoce la estructura particular de cada molécula individual la población puede estar constituida por moléculas con extremos idénticos (ll + rr), distintos (lr + rl) o una mezcla de ambas (ll + rr + lr + rl).

Los modelos de replicación elaborados para el genoma del virus de la vacuna (descritos en I.4) dan cuenta de la presencia de estructuras terminales en forma de horquilla conteniendo bases desapareadas que sufren inversiones de secuencia. Estos modelos son aplicables al caso del genoma del VPPA. En la figura 36 se ilustra la replicación de una molécula con extremos idénticos, según el modelo de autoiniciación, gracias al cual se generan moléculas hijas con todas las configuraciones posibles. Este mismo resultado se obtiene al replicar moléculas que presentan las otras tres posibles configuraciones.

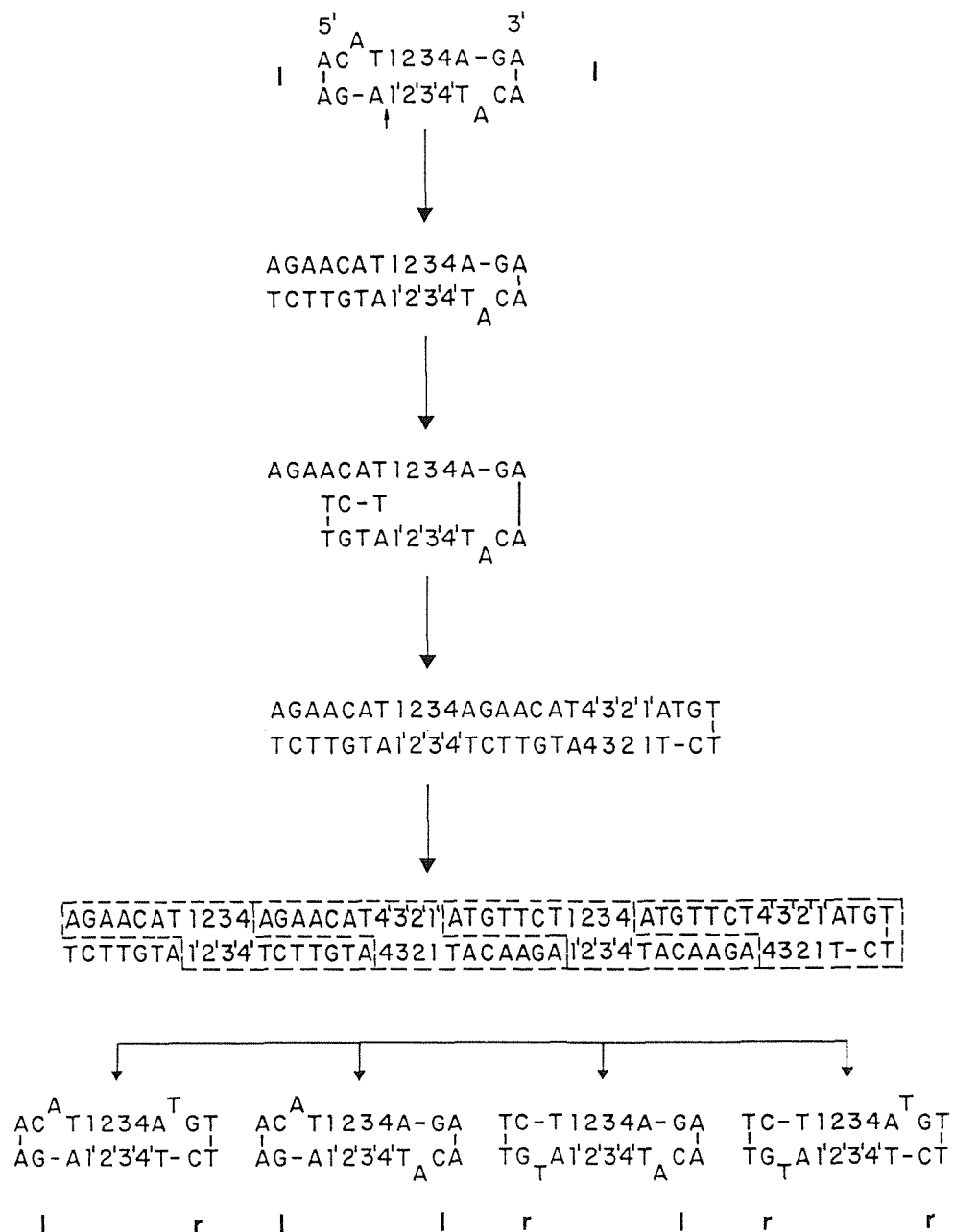


Figura 36. Generación de moléculas con todas las estructuras terminales posibles a partir de un único tipo de molécula parental. Se representa, con letras, la secuencia de una horquilla terminal hipotética de un tipo determinado (l) y con números, la secuencia de la parte central del genoma. Los símbolos ' denotan las secuencias complementarias. Si dicha molécula se replicara según el modelo de autoiniciación con una única interrupción en la cadena (figura 10a), se podría generar un tetrámero cuya escisión en secuencias específicas y ligación daría lugar a cuatro moléculas hijas con todas las configuraciones posibles de extremos: lr, ll, rl y rr.

El primer paso postulado en este modelo es la generación de una interrupción en la cadena polinucleotídica, con el fin de generar un extremo 3'-OH iniciador, en una localización específica de la repetición terminal invertida próxima a la horquilla terminal. Esto implica la existencia de una nucleasa específica, que reconociera una secuencia o estructura secundaria peculiares. La homología de secuencia detectada en las zonas adyacentes a la horquilla terminal del VPPA y del VV podría corresponder a una secuencia de reconocimiento utilizada por enzimas de estructura y función similares.

El modelo de replicación que implica la iniciación "de novo" (figura 3B), da lugar a dímeros circulares con orientación cabeza-cabeza, cola-cola como intermediarios replicativos. En el citoplasma de células infectadas por el VPPA se han detectado por hibridación fragmentos de restricción, que contienen las secuencias terminales, de doble longitud respecto de los fragmentos terminales del genoma del virión (157). Estos resultados demuestran la existencia de intermediarios replicativos de tamaño superior al monómero, en los que la orientación de los genomas unidad sería cabeza-cabeza, cola-cola.

Tanto el modelo de autoiniciación descrito anteriormente como el de iniciación "de novo" postulan este tipo de intermediarios. Por lo tanto cualquiera de ellos, o ambos a la vez, podría ser un mecanismo real de replicación del genoma del VPPA.

Los resultados experimentales descartan la existencia de intermediarios replicativos de tamaño superior al monómero en el que los genomas unidad tuviesen la orientación cabeza-cola, pero no la de otro tipo de intermediarios. En consecuencia, la replicación podría tener lugar, además de por aquel o aquellos mecanismos que dan lugar a los intermediarios detectados, por otros que den lugar a intermediarios de otro tipo. Un ejemplo de estos últimos sería el modelo de autoiniciación descrito en la figura 3A, que no da origen a concatémeros.

## 2. REPETICION TERMINAL INVERTIDA.

La secuenciación completa de la repetición terminal invertida presente en el extremo izquierdo ha permitido establecer su longitud en 2440 nucleótidos. Los experimentos previos de microscopía electrónica dieron como resultado una estimación de su longitud en  $2.09 \pm 0.21$  kpb (143), que se encuentra en concordancia con la longitud determinada por secuenciación.

El solapamiento de las lecturas individuales obtenidas a partir de cada subclon Bal 31 (IV.2.1) se ha basado en primera aproximación en el tamaño del inserto de los subclones, determinado por electroforesis en gel de agarosa. Sin embargo, la movilidad electroforética no permite una determinación precisa del tamaño de los fragmentos que contienen secuencias de la repetición terminal. Por ejemplo, los tamaños determinados por electroforesis de los fragmentos EcoRI-K' genómico, clonado y EcoRI<sub>1</sub>-HindIII son 4.8, 3.8 y 2.5 kpb, respectivamente, mientras que sus longitudes determinadas por secuenciación son 5.1, 4.2 y 3.0 kpb. Este hecho puede deberse a la peculiar estructura de la repetición terminal, compuesta en gran parte por repeticiones directas en tandem.

La propia secuencia determinada en cada lectura ha permitido el alineamiento exacto de las diversas lecturas para obtener la secuencia global. En el caso de las repeticiones directas en tandem, su carácter imperfecto ha permitido, te-

niendo en cuenta la secuencia de microheterogeneidades observada en cada lectura individual, alinearlas correctamente. Sin embargo, siempre cabe la posibilidad de que la zona común a dos lecturas o parte de ella, se encuentre en realidad duplicada de forma exacta en la secuencia.

Las dificultades que entraña la determinación de este tipo de secuencias se refleja en el hecho de que la estructura de los grupos de repeticiones en tandem presentes en las repeticiones terminales del virus de la vacuna (54) y del poxvirus bovino (66), no se ha determinado totalmente por secuenciación, sino integrando datos de secuenciación con mapas de restricción.

La organización general de los grupos de secuencias únicas y repetidas dentro de la repetición terminal invertida del VPPA es similar a la que exhiben las de los dos poxvirus estudiados, siendo los tamaños de los dos grupos de repeticiones directas más similares a los del poxvirus bovino (66).

Al comparar las secuencias consenso de las repeticiones en tandem del VPPA con las de estos dos poxvirus se observa una homología de secuencia elevada entre algunos segmentos de dichas secuencias, según se muestra en la figura 37. Un segmento de 19 nucleótidos de la secuencia de RDT27 presenta un grado de homología de un 75% con una zona de las repeticiones de 70 nucleótidos del virus de la vacuna. En el caso de la RDT34, un segmento de 17 nucleótidos presenta una homología del 80% con la repetición tipo B\* del poxvirus bovi-

RDT70-VV	AACTTTTTTACGACTCC	12/17
	■ ■ ■ ■ ■ ■ ■ ■ ■ ■	
RDT34-VPPA	ATTAAAATTATTACTACTGCTGTAGGCGTTAAAC	
	■ ■ ■ ■ ■ ■ ■ ■ ■ ■	
RDTB*-PB	AATATTTTTACGACT	12/15
RDT27-VPPA	CCATTTTATGGGCGTCAATCAATCACA	
	■ ■ ■ ■ ■ ■ ■ ■ ■ ■	
RDT54-VV	TTTTATGAG-ACCCATCAA	14/19

Figura 37. Homología de secuencia entre las secuencias de las RDT presentes en las RTI del VPPA y los poxvirus.

Se muestran las secuencias consenso completas de las RDT34 y RDT27 del VPPA, y las zonas homólogas presentes en las RDT de los poxvirus (VV, virus de la vacuna; PB, poxvirus bovino). Los puntos indican las bases idénticas con la secuencia de VPPA. En la parte derecha se indican el número de bases idénticas respecto a la longitud del segmento homólogo (54,66).

no (y menor con los otros tipos de repeticiones) y del 70% con la repetición de 54 nucleótidos del virus de la vacuna.

Esta homología sugiere un origen evolutivo común de ambos tipos de secuencia, o bien que se trata de un caso de evolución convergente.

El carácter restringido de la heterogeneidad de secuencia observada en las repeticiones en tandem (IV.4.2), puede deberse a que todas ellas procedan de una única secuencia por sucesivas duplicaciones génicas y evoluciones divergentes de cada secuencia. El análisis de la secuencia de las microheterogeneidades (figura 19) parece apoyar este argumento puesto que se observan grupos de RDT34 repetidos a su vez de forma directa (IV.4.2). Este es un rasgo en común con la estructura de la repetición terminal del poxvirus bovino (66).

El que las microheterogeneidades se restrinjan a unas pocas posiciones sugiere que la secuencia de las posiciones restantes quizás se haya conservado por medio de algún mecanismo específico.

La función de las RTI presentes en los genomas de poxvirus no se conoce, salvo en el caso del virus de la vacuna. La parte más interna de la RTI, que no contiene repeticiones en tandem, codifica cuatro polipéptidos (94). En el caso del VPPA, aunque existen dos fases de lectura abierta de pequeño tamaño presentes en la RTI no se detectan los polipéptidos correspondientes por experimentos de selección-traducción (57). Sin embargo, existe evidencia experimental



reciente de que la RTI se transcribe, correspondiendo probablemente a regiones 3'-no codificantes de RNAs mensajeros procedentes de la transcripción de las fases K'362, K'360 y D'363 (158).

La función de la RTI podría estar relacionada con su estructura peculiar: presencia de grupos de repeticiones en tandem, uno de ellos con un elevado número de subunidades, que presentan microheterogeneidad de secuencia en un reducido número de posiciones. Esta estructura podría favorecer interacciones específicas DNA-DNA o DNA-proteína necesarias durante el proceso de replicación.

Las interacciones DNA-DNA consistirían en el favorecimiento del reconocimiento de los extremos del genoma. En la figura 38 se muestran tres posibles variantes dentro del modelo de autoiniciación. Estas variantes surgen de la posibilidad de que se produjeran interrupciones específicas en uno o en ambos extremos del genoma, y en este último caso que la replicación se iniciase en uno solo o en ambos extremos 3'-OH resultantes. En el caso de dos interrupciones e iniciación solo a partir de una de ellas, la replicación de la cadena desplazada tendría lugar por medio de intermediario en forma de raqueta, posible gracias a la presencia de la RTI.

La presencia de las repeticiones en tandem favorecería inicialmente el proceso de apareamiento de las secuencias en RTI, mientras que la microheterogeneidad presente en las re-

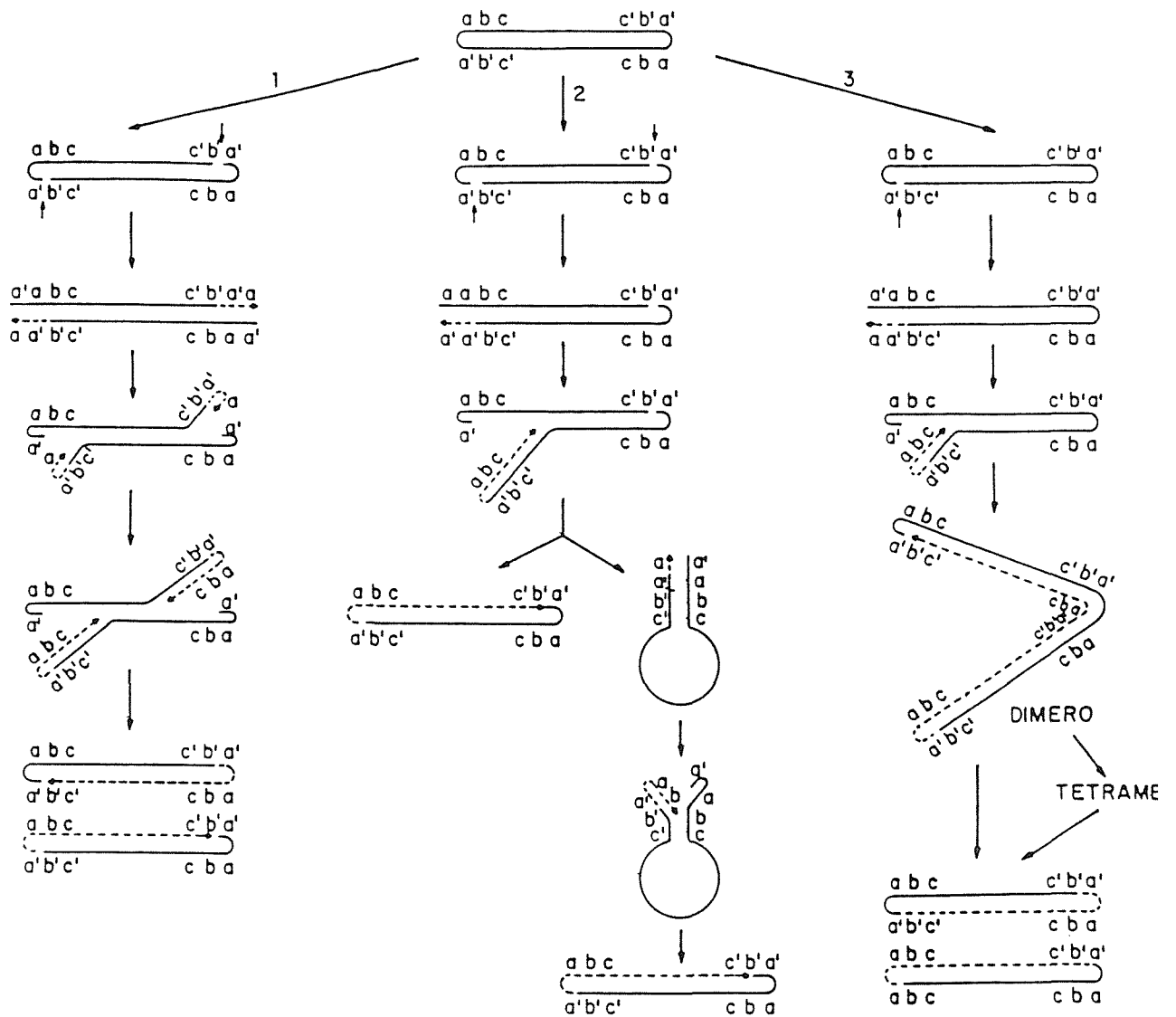


Figura 38. Modelos de replicación por autoiniciación de DNAs lineales con horquillas terminales. En el caso de generarse dos interrupciones en la cadena e iniciación solo a partir de una de ellas (2), se generaría un intermediario en forma de raqueta. Las secuencias de la RTI se simbolizan por  $abc$ ,  $a'b'c'$ .

peticiones aseguraría que el apareamiento fuese perfecto (65).

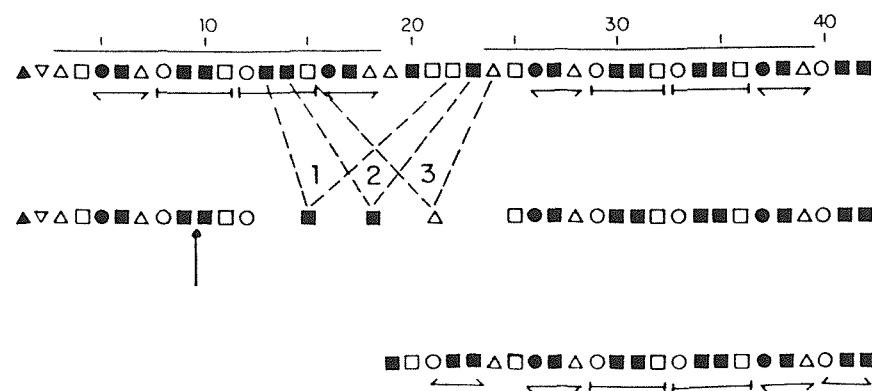
En cuanto a las interacciones DNA-proteína, las repeticiones en tandem podrían participar en el reconocimiento de la secuencia blanco por parte de la proteína encargada de generar interrupciones en la cadena polinucleotídica. Este reconocimiento se produciría en dos pasos, uno de unión débil pero específica a las repeticiones en tandem, tras lo cual la proteína se desplazaría a lo largo de la cadena polinucleotídica hasta encontrar el sitio de actuación, al que se uniría mas fuertemente. Las repeticiones en tandem aumentarían el tamaño de la secuencia que daría lugar a una interacción proteína-DNA productiva. Este modelo ha sido propuesto para el reconocimiento de los elementos transponibles FB de *Drosophila* por su transposasa (159) teniendo en cuenta el reconocimiento en dos pasos de la secuencia de actuación por parte del represor lac (160). El elemento FB4 contiene, dentro de su repetición terminal invertida, repeticiones en tandem (159).

El reconocimiento por parte de la proteína podría estar basado en la peculiar secuencia de las repeticiones en tandem, o bien en algun tipo de estructura terciaria peculiar inducida por la secuencia de las repeticiones. Esta estructura peculiar podría ser la responsable de la movilidad electroforética anómala de fragmentos que contienen estas repeticiones, comentada previamente.

Las diferencias observadas en las secuencias de las repeticiones terminales izquierda y derecha (figura 19 y tabla V) plantean la cuestión de si en una molécula las repeticiones terminales son perfectas o no. Puesto que al clonar los fragmentos terminales se han amplificado las secuencias pertenecientes, muy probablemente, a extremos de diferentes genomas individuales, la heterogeneidad de secuencia observada puede ser intermolecular. Debido a que no se dispone de la secuencia completa de la repetición terminal derecha, no se puede asegurar que las diferencias sean únicamente las detectadas. No se puede descartar tampoco que exista heterogeneidad, tanto intra como intermolecular, en el número de RDT34 presentes en la repetición, en todo caso de reducida extensión, puesto que ni el mapeo de restricción (52) ni la visualización de heterodúplex por microscopía electrónica (143) la han puesto de manifiesto. En el caso del poxvirus bovino, el análisis de la repetición terminal derecha (no secuenciada) por mapeo fino de restricción no detecta diferencias de secuencia entre ambas repeticiones (66).

Otra posibilidad es que las RTI posean una homología de secuencia perfecta, y que las diferencias detectadas tengan su origen en un artefacto producido en el proceso de clonaje y amplificación del fragmento terminal derecho. Las diferencias observadas se localizan exclusivamente en la zona de las repeticiones en tandem, y dentro de ellas, en las más terminales. La inestabilidad replicativa de las secuencias repetidas es un hecho descrito para fragmentos clonados

a)



b)

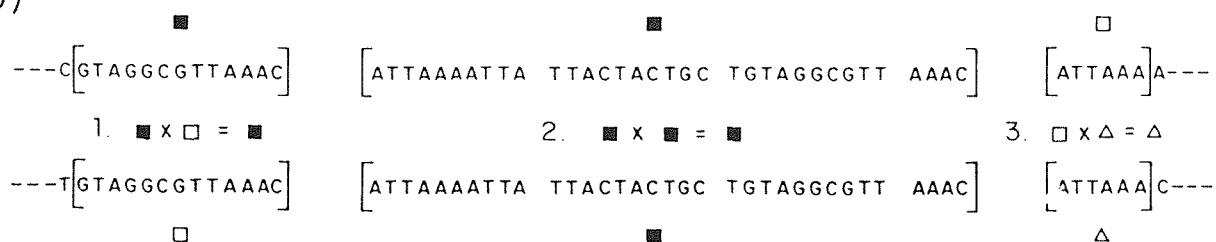


Figura 39. Mecanismo de generación de la RTI presente en el fragmento terminal derecho clonado. a) A partir de una secuencia idéntica a la existente en el fragmento terminal izquierdo, la delección introducida por la nucleasa S1 se extendería hasta la RDT34 señalada por la flecha. Un proceso de recombinación intra o intermolecular entre las RDT34 unidas por trazos discontinuos daría lugar a la estructura observada en el fragmento terminal derecho. Esta se muestra en la tercera línea. Se ha originado otra delección de nueve RDT34 y aparecen RDT34 de distinto tipo en posiciones equivalentes de la RTI. b) Segmento en el que una recombinación homóloga daría lugar a la estructura observada. Incluye una RDT34 completa y parte de las RDT34 adyacentes, debido a la secuencia de microheterogeneidades.

	POXVIRUS	VPPA	IRIDOVIRUS
ESTRUCTURA DEL DNA	Secuencia única con horquillas y repeticiones invertidas terminales		Secuencia permutada circularmente con repeticiones directas terminales
ESTRUCTURA DEL VIRUS	Simetría compleja Glicoproteína estructural	Simetría icosaédrica Sin glicoproteína estructural	
SINTESIS DEL RNA	El virus contiene los enzimas necesarios para la síntesis de mRNA La síntesis de mRNA es independiente de la RNA polimerasa II celular Los mRNA están poliadenilados		El virus no contiene los enzimas necesarios para la síntesis de mRNA La síntesis de mRNA es dependiente de la RNA polimerasa II celular Los mRNA no están poliadenilados
REPLICACION DEL DNA	La replicación del DNA es independiente del núcleo celular	La replicación del DNA es dependiente del núcleo celular	
MORFOGENESIS EN PRESENCIA DE $\alpha$ -AMANITINA	Virus inmaduro	Virus infectivo	No virus

Figura 40. Comparación de las propiedades del VPPA , los poxvirus y los iridovirus.

utilizando *E. coli* como bacteria huésped (116). La secuencia presente en el clon correspondería al producto de una alteración causada por la replicación en *E. coli* que ya es susceptible de una replicación fiel.

En la figura 39 se muestra un mecanismo hipotético de generación de la secuencia presente en el clon. La nucleasa S1 habría producido una delección terminal mucho menor que la observada, pero que originaría que el fragmento ligado al vector tuviera en un extremo una RDT34, y no secuencia única como en el caso del fragmento terminal izquierdo clonado. Este hecho conferiría a la zona del inserto inestabilidad replicativa. Una recombinación homóloga intramolecular podría tener lugar entre las RDT34 señaladas en la figura 39, con lo que se obtendría la secuencia presente en el clon. La zona en la que una recombinación produciría la secuencia de microheterogeneidad observada incluye aproximadamente 2 RDT34 (46 nucleótidos).

Las características estructurales descritas (horquilla terminal y repetición invertida terminal) hacen el genoma del VPPA más similar al de los poxvirus que al de los iridovirus, grupo en el que se encuentra clasificado por las características morfológicas de la partícula viral. En la figura 40 se resumen las características del VPPA comparadas con las de ambos grupos de virus. Sus características únicas sugieren la conveniencia de clasificar al VPPA en un grupo distinto, del que, por ahora, sería el único representante conocido (161).

### 3. SECUENCIAS CODIFICANTES.

La secuenciación de las zonas adyacentes a las repeticiones terminales y el análisis de la secuencia obtenida han permitido localizar varias fases de lectura abierta (IV.5). Tres de ellas, las denominadas K'362, K'360 y D'363, son de un tamaño casi idéntico y presentan un elevado grado de homología de secuencia, tanto a nivel de aminoácidos como de nucleótidos (figuras 25 y 26). Una cuarta de tamaño mucho menor (D'42), presenta asimismo un alto grado de homología con las anteriores en sus primeros 30 codones (figura 25). La presencia de estas secuencias homólogas próximas a la repetición terminal da cuenta de las repeticiones invertidas internas detectadas en los fragmentos terminales por técnicas de microscopía electrónica (I.4).

Los tamaños moleculares, predichos a partir de los datos de secuencia, de las proteínas que se sintetizarían en el caso de que estas fases de lectura abierta se expresaran están en concordancia con los de algunos polipéptidos mapeados en los fragmentos terminales por experimentos de selección-traducción (figura 2 y tablas VI y X). Las proteínas de 41.5 y 42 K, seleccionadas por los fragmentos EcoRI-K' y D', corresponderían a la expresión de las fases K'362, K'360 y D'363; la de 17 K, seleccionada por los fragmentos EcoRI-K', podría corresponder a la fase K'L177.



Existe asimismo evidencia experimental directa de que las fases K'362 y K'360 y D'363 se transcriben, obtenida mediante hibridación de DNA a RNAs mensajeros virales separados electroforéticamente y transferidos a filtro de nitrocelulosa (158).

Se han localizado de forma precisa los sitios de iniciación de la transcripción de las fases K'362, K'360 y D'363. Las secuencias adyacentes en posición 5' al inicio de la transcripción no contienen señales reguladoras eucarióticas canónicas. Sin embargo, presentan segmentos homólogos a secuencias comúnmente encontradas en la región adyacente al inicio de la transcripción de varios genes tempranos del VV (I.4). Estas secuencias se muestran en la figura 41.

Experimentos recientes han demostrado que la RNA polimerasa codificada por el VPPA es capaz de reconocer eficientemente "in vivo" la región promotora del gen 7.5 K del virus de la vacuna (162). Estos datos sugieren que las RNA polimerasas codificadas por ambos virus reconocen señales reguladoras homólogas.

La presencia de secuencias homólogas a la consenso de unión de RNAs mensajeros al ribosoma eucariótico, adyacentes en posición 5' al ATG iniciador de todas las fases detectadas (figura 23), indica que los productos resultantes de la transcripción de dichas fases resultan probablemente traducidos de forma eficiente por la maquinaria celular.

El análisis de las proteínas, tanto presentes en la partícula viral como inducidas por la infección, no revelaba

presencia de ninguna proteína que se sintetice en los primeros estadios de la infección de tamaño molecular similar al predicho a partir de la secuencia de nucleótidos (28,163). Puesto que la detección de polipéptidos de tamaño concordante con los datos de secuencia se ha realizado utilizando sistemas de transcripción-traducción "in vitro", existe la posibilidad de que las proteínas sintetizadas "in vivo" sufran modificación postraducciona, con lo que su identificación debería realizarse por otros criterios distintos del tamaño molecular.

Una aproximación consistiría en la generación de antisueros monoespecíficos o anticuerpos monoclonales que reconociesen epítopos de dichas proteínas. El antígeno podía consistir en proteína sintetizada en células procariontes por técnicas de ingeniería genética, o bien en péptidos sintéticos. La elección de los péptidos a sintetizar se basaría en las predicciones de hidrofiliidad, antigenicidad potencial y estructura secundaria de giro  $\beta$  (figuras 28-29 y 33-34).

### 3.1. FAMILIA 360.

La homología de secuencia que exhiben las fases K'362, K'360 y D'363 se refleja en la similitud de sus composiciones de aminoácidos y frecuencias de usos de codones (tablas VI y VIII). La frecuencia de uso de codones de estas fases parece estar muy condicionada por el alto contenido en A+T de la secuencia. En la tabla XII se comparan dichas frecuen-

<u>TAATTTATCCAATCTGTATTCTTGAAAACATTAAAACTTAAACCATTTTGC</u> <u>A</u>	K'362	
GAGATTTATTAT <u>TATAATTTACAGGCTTTATTAAAAAAAATAGAATTTAGTA</u> <u>T</u>	K'360	
ATGACTAAAAT <u>CATATTATAATGCCGTGCAAAAAATAATTATTTTTCGGTTAA</u> <u>A</u>	D'363	
AATAATTA	GAAAATATA	7.5 K
TATTATTA	AATAATATA	19 K
TATAATAA	AAAAAAAAG	42 K
TATATTAA	AATAATTAA	TK

Figura 41. Secuencias adyacentes en posición 5' al inicio de la transcripción de las fases K'362, K'360 y D'363.

La base recuadrada corresponde en cada caso al inicio de la transcripción. Los segmentos subrayados presentan homología con secuencias localizadas en zonas análogas de genes tempranos del VV, las cuales se muestran en la parte inferior de la figura (87,88,95).

TABLA XII. Frecuencia de uso de codones porcentual para cada aminoácido.

	Codon	360	K'L177	Mam		Codon	360	K'L177	Mam
Arg	CGA	10	20	5	Gly	GGA	20	55	15
	CGC	0	0	18		GGC	6	0	42
	CGG	24	20	16		GGG	33	0	18
	CGT	14	40	13		GGT	41	45	26
Leu	AGA	31	20	21	Val	GTA	44	36	6
	CTA	14	10	8		GTC	14	18	25
	CTC	1	0	24		GTG	20	9	57
	CTG	5	0	53		GTT	22	36	12
Ser	CTT	15	30	7	Lys	AAA	74	72	27
	TTA	48	40	3		AAG	26	28	73
	TTG	17	20	7	Asn	AAC	31	0	75
	TCA	21	17	12		AAT	69	100	25
Thr	TCC	6	17	22	Gln	CAA	63	80	23
	TCG	5	0	4		CAG	37	20	77
	TCT	30	25	20	His	CAC	13	25	71
	AGC	12	25	27		CAT	87	75	29
Pro	AGT	26	17	16	Glu	CAA	71	60	38
	ACA	24	57	19		CAG	29	40	62
	ACC	19	9	44	Asp	GAC	29	27	59
	ACG	15	9	10		GAT	71	73	41
Ala	ACT	43	25	27	Tyr	TAC	8	13	63
	CCA	55	50	23		TAT	92	87	37
	CCC	10	20	38	Cys	TGC	24	20	58
	CCG	15	10	13		TGT	76	80	42
	CCT	20	20	27	Phe	TTC	28	17	68
	GCA	28	17	14		TTT	72	83	32
	GCC	19	33	46	Ile	ATA	48	33	13
	GCG	12	0	9		ATC	15	7	53
	GCT	41	50	31		ATT	37	60	34

Los valores en cada columna corresponden a: 360, la frecuencia porcentual global de las fases K'362, K'360 y D'363; K'L177, frecuencia porcentual de dicha fase; Mam, frecuencia porcentual de un conjunto de referencia de 26 genes de mamíferos (167).

cias con las promedio de un conjunto de genes eucarióticos (164). La frecuencia de aparición de A o T en la tercera base de los codones es muy superior a la existente en estos genes de referencia.

Las composiciones de aminoácidos de las tres fases son bastante similares. Los pI teóricos de K'362 y D'363 tienen valores próximos y se encuentran en el rango ácido, mientras que el correspondiente a K'360 es muy distinto y claramente básico. Este dato contrasta con el hecho de que el par de secuencias con mayor grado de homología global son K'362 y K'360 (figura 25).

La homología entre las secuencias de la familia 360 se concentra en algunos segmentos de las mismas. Los segmentos denominados III, V y X en la figura 25 son especialmente homólogos tanto si se comparan secuencias de nucleótidos como de aminoácidos (tabla VIII). De estos tres segmentos, el V y el X presentan una acumulación de residuos ácidos.

Esta homología de secuencia global se refleja en los resultados de los análisis predictivos de hidropatía y estructura secundaria (figuras 27 y 28). La asignación a las zonas análogas de las tres secuencias de un determinado tipo de estructura secundaria o grado de exposición al solvente es similar en las tres fases.

Las predicciones de participación de determinados segmentos de las secuencias en epitopos continuos arrojan resultados diferentes para cada fase. Las zonas de mayor hidrofiliidad relativa y que presentan valores de antigenici-

dad potencial superiores a 0.05 se localizan en posiciones diferentes en las tres secuencias (figura 29). En esas posiciones la homología de secuencia de aminoácidos es relativamente baja. Si se utilizaran como antígenos péptidos sintéticos conteniendo las secuencias correspondientes a estas posiciones, los anticuerpos inducidos podrían resultar específicos de una sola de las proteínas de la familia.

Los sitios potenciales de N-glicosilación presentes en las tres fases no se localizan en posiciones análogas de las tres secuencias, como se observa en la figura 28. Los análisis predictivos de exposición al solvente y participación en epitopos continuos no son positivos en ninguna de estas posiciones, según se muestra en la tabla IX. No parece por lo tanto probable que las proteínas estuviesen N-glicosiladas.

La homología detectada entre distintas zonas de una misma secuencia aparece centrada en residuos de cisteína (figura 30). Estos residuos juegan frecuentemente un importante papel en el mantenimiento de la estructura terciaria de las proteínas. Entre las dos primeras cisteínas y las dos segundas (posiciones 105 y 137; 256 y 282 de K'362, respectivamente) existe una distancia similar de aproximadamente 30 residuos.

Si las mitades N- y C-terminales de la secuencia se hubieran originado por un proceso de duplicación, es posible que estas zonas hayan mantenido un cierto grado de homología debido a una cierta similaridad estructural persistente en-

tre los segmentos duplicados con participación de los residuos de cisteína.

No existen datos experimentales acerca de la función de las supuestas proteínas de la familia 360. Los resultados de los análisis predictivos de estructura secundaria e hidropatía permiten postular que la estructura tridimensional de las tres proteínas sería similar. Por extrapolación, se podría esperar que sus funciones también serían similares. Esta hipótesis debería tener en cuenta sin embargo las diferencias que las preedicciones de otras características de las proteínas arrojan resultados distintos: pI y participación de determinadas zonas de la secuencia en epitopos continuos.

En el genoma del VPPA existen otras familias de genes homólogos en tandem localizados por secuenciación de la zona del genoma correspondiente a los fragmentos de restricción EcoRI-L, U', V y X (figura 1). Existe evidencia experimental de que al menos una de ellas es funcional en cuanto a su transcripción (165).

La localización de todas estas familias de genes en una zona del genoma, el extremo izquierdo, en la que se observa asimismo variabilidad genética entre diversos aislados virales (166), podría sugerir que la función de las proteínas codificadas sería su participación en un mecanismo de escape del virus al sistema inmune del huésped por variabilidad antigénica. En el caso de las fases K'362 y K'360, existe un aislado viral que presenta una fase de lectura abierta que

contiene la parte N-terminal de K'362 y la C-terminal de K'360 (166) generada probablemente por un mecanismo de recombinación homóloga.

Sin embargo, la zona del genoma de mayor variabilidad genética, determinada a partir de la comparación de mapas de restricción, se encuentra localizada más hacia el interior del genoma, entre 8 y 11 kpb a partir del extremo izquierdo. Es en esta zona hipervariable en la que se localizan las otras dos familias de genes homólogos en tandem. Especialmente, la familia con cinco copias detectadas hasta el momento podría estar implicada en el escape del sistema inmune por cambio antigénico en la superficie del virus, ya que existe una proteína viral de 12 K identificada como externa por accesibilidad a detergentes y microscopía electrónica, que se une específicamente a células VERO (167). Esta proteína, por su tamaño molecular, podría ser la codificada por dichos genes.

La existencia de un aislado viral con una estructura de la familia 360 distinta de la presente en el virus adaptado a células VERO implica que la existencia de las tres copias génicas no es indispensable para la supervivencia viral y apoya asimismo la hipótesis de una función similar para las tres proteínas.

En lo que respecta al origen de esta familia de genes, el análisis de los datos de comparación de secuencias (figura 26) permite postular que las fases K'362 y K'360 se han originado por evolución divergente tras la duplicación génica.



ca de una fase ancestral. Esta fase ancestral localizada en el extremo izquierdo y otra, localizada en el extremo derecho y precursora de la actual D'363, procederían de un ancestro común, localizado en uno cualquiera de los dos extremos. Aunque el mecanismo preciso de la duplicación de secuencias de un extremo a otro del genoma no se conoce, existen numerosos ejemplos de tales fenómenos en los genomas de los poxvirus (168).

El origen de la fase D'42 dada su gran similaridad, en la parte N-terminal con la K'362 (figura 26) no parece poder explicarse por ninguno de los dos mecanismos descritos, si tenemos en cuenta su localización: en el extremo derecho, como D'363, pero próxima a la repetición terminal, como K'360 en el extremo izquierdo. Una recombinación no homóloga entre distintos extremos de dos genomas que contuvieron precursores de las fases K'360, K'362 y D'363, podría dar cuenta de la aparición de D'42.

De todas las familias multigénicas descritas en vertebrados, solo la correspondiente a los fibrinopéptidos humanos (169) presenta copias con transcripción en distinta orientación, como es el caso de la familia 360 de VPPA. Los casos descritos para familias multigénicas de invertebrados son mucho más abundantes (170). La presencia de miembros con pI tan dispares dentro de una familia de proteínas de función similar sería otro rasgo destacable, de verificarse en las proteínas nativas.

Sin embargo, la orientación de las copias de la familia 360 resulta menos singular si la comparamos con las proteínas codificadas en la RTI del virus de la vacuna, las cuales, debido al carácter invertido de la repetición, se transcriben asimismo con orientaciones opuestas.

Dada la homología existente entre las secuencias de la familia 360, se podría considerar que los extremos del genoma del VPPA constituyen una repetición terminal invertida que incluye zonas de homología perfecta e imperfecta. Las primeras corresponderían a la RTI propiamente dicha, compuesta mayoritariamente por repeticiones directas en tandem, y las segundas, a las secuencias codificantes de la familia 360. Desde este punto de vista, la estructura de los extremos de los genomas del VPPA y del VV resultan muy similares.

Además del caso del VPPA, existen dos grupos de fases de lectura abierta que codificarían proteínas homólogas en el genoma de otro deoxivirus animal, el virus de Epstein-Barr, cuya secuencia de nucleótidos se ha determinado completamente (104). Se trata de tres fases contiguas denominadas BERF1, BERF2b y BERF4, de tamaños 92, 93 y 96 K, respectivamente, sobre las que no existen datos acerca de su expresión ni de la función de sus posibles productos génicos.

Las VPg del virus de la fiebre aftosa, presentes en tres copias homólogas en tandem, constituyen el único caso descrito de familias de genes virales que codifiquen proteínas de función conocida (171).

### 3.2. K'L177.

Las características de esta posible proteína son presumiblemente muy distintas de las que presentan las pertenecientes a la familia 360. Su pI teórico es aproximadamente neutro (tabla X). Presenta un perfil de hidropatía con valores mucho menos extremados, si se exceptúa la zona N-terminal que resulta predicha como péptido señal (figura 32). El resto de la secuencia resulta predicha como moderadamente expuesta al solvente.

La estructura secundaria predicha presenta abundante proporción de estructura aperiódica y de giro  $\beta$  (figura 33). Sus residuos de cisteína se encuentran agrupados en dos zonas. La predicción de participación de determinados grupos de residuos en epitopos continuos no es coincidente dependiendo del método utilizado salvo en dos zonas (figura 34).

Sin embargo, su frecuencia de uso de codones parece sufrir la influencia del elevado contenido en A+T de la secuencia codificante. Como se observa en la tabla XII, su distribución exhibe la misma tendencia a restringir el uso de G o C en tercera posición, como es el caso en las secuencias de la familia 360.

La función de esta proteína podría estar relacionada con la presencia de péptido señal en su extremo amino. Quizá se encuentre anclada en alguna de las envueltas lipídicas de la partícula viral.

#### 4. CONCLUSIONES.

1.- La unión covalente terminal del genoma del VPPA está constituida por una horquilla polinucleotídica, con lo que la molécula de DNA bicatenario consiste en una única cadena in-interrumpida.

2.- La horquilla terminal del genoma del VPPA existe en dos formas, presentes en los dos extremos. Sus secuencias son invertidas y complementarias, y su estructura secundaria contiene bases desapareadas.

3.- La repetición terminal invertida del genoma del VPPA contiene a su vez repeticiones directas y directas en tandem separadas por zonas de secuencia única.

4.- Estas repeticiones directas son imperfectas. Sin embargo, las diferencias se localizan en un reducido número de posiciones dentro de cada tipo de repetición.

5.- En la región adyacente a la RTI existen varias fases de lectura abierta. Tres de ellas tiene aproximadamente la misma longitud (360 codones) y presentan un elevado grado de homología de secuencia. Dos de ellas se localizan en el extremo izquierdo y la tercera en el derecho.

6.- Una cuarta fase de tan solo 42 codones de longitud localizada en el extremo derecho presenta una elevada homología en sus primeros 30 tripletes con los codones iniciales de las tres fases anteriores.

## BIBLIOGRAFIA

1. Dekock, G., Robinson, E.M. y Keppel, J.J.G. (1940). Swine fever in South Africa. *Onderstepoort Jour. Vet. Sci and Anim. Inv.* 14, 31-93.
2. Montgomery, R.E. (1921). On a form of swine fever occurring in British East Africa (Kenya Colony). *Jour Comp. Path. and therap.* 34, 159-191, 243-262.
3. Maurer, F.D., Griesemer, R.A. (1970). African Swine Fever. *Diseases of Swine*. H.W. Dunne. The Iowa State University Press.
4. Manso Ribeiro, J., Rosa Azevedo, J.A., Texeira, M.J.O., Braco Forte, M.C., Rodriguez Ribeiro, A.M., Oliveira, E., Noronha, F., Grave Pereira, C., y Dias Vigario, J. (1958). An atypical strain of swine fever virus in Portugal. *Bull. Off. Int. Epiz.* 50, 516-534.
5. Lucas, A. y Haag, J., Larenaudie, B. (1971). Peste porcina africana. Editorial Acridia.
6. De Tray, D.E. (1963). African swine fever. *Advan. Vet. Sci.* 8, 299.
7. Coggins, L. (1974). African swine fever virus. Pathogenesis. *Progr. Med. Virol.* 18, 48.
8. Cowan, K.M. (1961). Immunological studies on ASFV. I. Elimination of the procomplementary activity of swine serum with formalin. *J. Immunol.* 86, 465-470.
9. Boulanger, P. Bannister, G.L., Gray, D.P., Ruckenbauer, G.M. y Willis, N.G. (1967). African swine fever. III. The use of agar double-diffusion precipitation test for the detection of the virus in swine tissues. *Can. J. Comp. Med.* 31, 7-11.

10. Stone, S. S., De Lay, P.D. y Sharmen, E.L. (1968). The antibody response in pigs inoculated with attenuated ASFV. *Can. J. Comp. Med.* 32, 455-460.
11. Malmquist, W.A. (1963). Serologic and immunologic studies with African swine fever virus. *Am. J. Vet. Res.* 24, 450.
12. Pan, I.C., Trautman, R., De Boer, C.I. Hess, W. R. y Tessler, J. (1974). African swine fever detection of antibody by reverse single radial immunodiffusion. *Am. J. Vet. Res.* 35, 629-631.
13. De Boer, C.I. (1967). Studies to determine neutralizing antibody in sera from animals recovered from ASF and laboratory animals inoculated with African virus with adjuvants. *Arch. Gesamte virusforsch.* 20, 164-179.
14. Shimizu, M, Pan, I.C. y Hess, W.R. (1977). Cellular immunity demonstrated in pigs infected with African swine fever virus. *Am. J. Vet. Res.* 38, 27-31.
15. Mossey, R.J. y Shochetman, G. (1981). Viral epitopes and monoclonal antibodies: Isolation of blocking antibodies that inhibit virus neutralization. *Science* 213, 447-449.
16. Enjuanes, L., Sanz, A., Barreno, B. G., Nogal, M. y Viñuela, E. Partial neutralization of ASFV by monoclonal and polyclonal antibodies. *Man. en prep.*
17. Sanz, A., Garcia-Barreno, B., Nogal, M.L., Viñuela, E. y Enjuanes, L. (1985). Monoclonal antibodies specific for african swine fever virus proteins. *J. Virol.* 54, 199-206.
18. Cuillel, M., Tripier, F., Braunwald, J. y Sacrot, B. (1979). A low resolution structure of frog virus 3. *Virology* 99, 277-285.



19. Wrigley, N.G. (1970). An electron microscope study of the structure of Tipula iridescent virus. J. Gen. Virol. 6, 169-173.
20. Breese S.S. y DeBoer, C.J. (1966). Electron microscope observation of ASFV in tissue culture cells. Virology 28, 420-428.
21. Granoff, A. (1969). Viruses of Amphibia. Curr. Top. Microbiol. Immunol. 50, 107-137.
22. Tripier, F., Markovic, L., Braunwald, J. y Kim, A. (1975). Dounees nouvelles sur la structure et le bourgeonnement du FV3. Ann. Microbiol. (Paris) 126B, 447-460.
23. Carrascosa, J.L., Carazo, J.M., Carrascosa, A.L., Garcia, N., Santisteban, A. y Viñuela, E. (1984). General morphology and capsid fine structure of ASFV particles. Virology 132, 160-172.
24. Tripier, F. y Kim, A. (1973). Mise en evidence de L'ultra-structure du FV3. Ann. Microbiol. (Paris). 124A, 155-168.
25. Spear, P.G. y Roizman, B. (1972) Proteins specified by herpes simplex virus. V. Purification and structural proteins of the herpesvirion. J. Virol. 9, 143-159.
26. Hess, W.R. (1971). African swine fever virus, en Virology Monographs (Gard, S., Hallauer y C. Meyer K.F. eds). Springer Verlag, New York, vol. 9, pp 1-33.
27. Trautman, R., Pan, I.C. y Hess W. R. (1980). Sedimentation coefficient of African swine fever virus. Amer. J. Vet. Res. 41, 1874-1878.
28. Carrascosa, A.L., del Val, M., Santarén, J.F. y Viñuela, E. (1985). Purification and properties of African swine fever virus J. Virol. 54, 337-344.

29. Breese S.S. Jr y de Boer C. J (1966) Electron microscope observations of African swine fever virus in tissue culture cells. *Virology* 28, 420-428.
30. Haag, J., Lucas A., Larenaudie, B., Ruiz Gonzalvo, F. R. y Carnero R. (1966). Peste porcine africaine. Recherches sur la taille et la morphologie du virus. *Rec. Med. Vet.* 142, 801-808.
31. Malmquist, W. A. y Hay, D. (1960) Hemadsorption and cytopathic effect produced by African swine fever virus in swine bone marrow and buffy coat cultures. *Amer. J. Vet. Res.* 21, 104-108.
32. Carrascosa, A. L., Santaren, J. F. y Viñuela, E. (1982). Production and titration of African swine fever virus in porcine alveolar macrophages. *J. Virol. Methods* 33, 303-310.
33. Polatnick, J., Pan I. C. y Gravell, M. (1974). Protein Kinase activity in African swine fever virus. *Arch. ges. Virusforsch.* 44, 156-159.
34. Black, D. N. y Brown, F. (1976). Purification and physicochemical characteristics of African swine fever virus. *J. gen. virol.* 32, 509-518.
35. Tabarés, E., Gonzalvo, F.R., Marcotegui, M.A. y Ordas, A. (1977). Growth, purification and fractionation of African swine fever virus. *Comm. Eur. Comm.* EUR5904 EN, pp 507-531.
36. Vigario, J. D., Castro Portugal, F. L., Ferreira, C. A. y Festas, M. B. (1977). Purification and study of the structural polypeptides of African swine fever virus. *Comm. Eur. Comm.* EUR 5904 EN, pp 469-482.
37. Kuznar, J., Salas, M. L. y Viñuela, E. (1980). DNA-dependent RNA polymerase in African swine fever virus. *Virology* 101, 169-175.

38. Tabarés, E., Martinez, E. y Escribano, J. M. (1983). Proteins specified by African swine fever virus. IV. Glycoproteins and phosphoproteins. Arch. Virol. 77, 167-180.
39. Tabarés, E., Marcotegui, M., Fernandes, M. y Sánchez Botija, C. (1980). Proteins specified by African swine fever virus. I. Analysis of viral structural proteins and antigenic properties. Arch. Virol. 66, 107-117.
40. del Val, M., Carrascosa, J. L. y Viñuela, E. (1986). Glycosylated components of ASFV particles. Virology, en prensa.
41. Kates, J.R. y Mc. Auslan, B.R. (1967). Pox-virus DNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 58, 134-141.
42. Munyon, W., Paoletti, E. y Grace, J.T., Jr. (1967). RNA polymerase activity in purified infections vaccinia virus. Proc. Natl. Acad. Sci. U.S.A. 58, 2280-2287.
43. Salas, M.L., Kuznar, J. y Viñuela, E. (1981). Polyadenylation, methylation and capping of the RNA synthesized in vitro by African swine fever virus. Virology 113, 484-491.
44. Kuznar, J., Salas, M. L. y Viñuela, E. (1981). Nucleoside triphosphate phosphohydrolase activities in African swine fever virus. Arch. Virol. 96, 307-320.
45. Salas, M. L., Kuznar, J. y Viñuela, E. (1983). Effect of rifamycin derivatives and coumermycin A1 on in vitro RNA synthesis by African swine fever virus. Arch. Virol. 77, 77-80.
46. Moss, B. (1974) en Reproduction of poxviruses. (Fraenkel-Conrat, H. y Wagner, R.R. eds) Plenum, New York. Comprehensive Virology, vol. 3, pp 405-474.

47. Dales, A. y Pogo, B.G.T. (1981) en *Biology of poxviruses* (Kingsbury, D.W. y Zur Hansen, H. eds). Springer, Berlin-Heidelberg-New York .Virology monographs, vol 8 pp. 1-109.
48. Haag, J. Larenaudie, B. y Gonzalvo F. R. (1965). Peste porcine africaine. Action de la 5-iodo-2'-desoxyuridine sur la culture du virus in vitro. *Bull. off. Int. Epiz.* 63, 717-722.
49. Plowright, W., Brown, F. y Parker, J. (1966). Evidence for the type of nucleic acid in African swine fever virus. *Arch. ges virusforsch.* 19, 289-304.
50. Enjuanes, L., Carrascosa, A.L. y Viñuela, E. (1976) Isolation and properties of the DNA of African swine fever virus. *J. Gen Virol.* 32, 479-492.
51. Ortín, J., Enjuanes, L. y Viñuela, E. (1979) Crosslinks in African swine fever virus DNA. *J. Virol.* 31 579-583.
52. Almendral, J.M., Blasco, R., Ley, V., Beloso, A., Talavera, A. y Viñuela, E. (1984). Restriction site map of African swine fever virus DNA *Virology* 133, 258-270.
53. Geshelin, P. y Berns, K. I. (1974). Characterization and localization of the naturally occurring crosslinks in vaccinia virus DNA. *J. Mol. Biol.* 88, 785-796.
54. Baroudy, B. M., Venkatesan, S. y Moss, B. (1982). Incompletely base-paired flip-flop terminal loops link the two strands of the W genome into one uninterrupted polynucleotide chain. *Cell* 28, 315-324.
55. Ley, V., Almendral, J.M., Carbonero, P., Beloso, A., Viñuela, E. y Talavera A., (1984). Molecular cloning of African swine fever virus DNA. *Virology* 133, 249-257.

56. Alcamí, A. y Carrascosa, A.L. (1985). Adsorción y penetración del VPPA en células VERO. Seminario Hispano-Norteamericano sobre peste porcina africana.
57. Salas, M.L., Rey-Campos, J., Almendral J. M., Talavera, A. y Viñuela, E.(1986). Transcription and translation maps of ASFV.Virology, en prensa.
58. Salas, J., com. pers.
59. Polatnick, J. y Hess, W. R. (1972). Increased deoxyribonucleic acid polymerase activity in African swine fever virus infected cells. Arch. ges. Virusforsch. 38, 383-385.
60. Moreno, M. A., Carrascosa, A. L., Ortin, J. y Viñuela, E., (1978). Inhibition of African swine fever virus replication by phosphonoacetic acid. J. gen. Virol. 93, 253-258.
61. Ortin, J. y Viñuela. E. (1977). Requirement of cell nucleus for African swine fever virus replication in VERO cells. J. Virol. 21, 902-905.
62. Pan, I. C., Shimizu, M. y Hess, W. R (1980). Replication of African swine fever virus in cell cultures. Amer. J. Vet. Res. 41, 1357-1367.
63. Mackett, M. y Archard, L.C. (1979). Conservation and variation in orthopoxvirus genome structure. J. Gen. Virol 45, 683-701.
64. Dumbell, K.R. y Archard L. C. (1980). Comparison of white pock (h) mutants of monkeypox virus with parental monkeypox and with variola-like viruses isolated from animals. Nature 286, 29-32.
65. Wittek, R. y Moss, B. (1980). Tandem repeats within the inverted terminal repetition of vaccinia virus DNA. Cell 21, 277-284.

66. Pickup, D.J., Bastia, D., Stone, H.O. y Joklik, W.K. (1982) Sequence of terminal regions of cowpox virus DNA. Arrangement of repeated and unique sequence elements Proc. Natl. Acad. Sci. USA 79, 7112-7116.
67. Baroudy, B.M. y Moss, B. (1982) Sequence homologies of diverse tandem repetitions near the ends of vaccinia virus genome suggest unequal crossing over Nucl. Ac. Res. 10, 5673-5679.
68. Baroudy, B.M., Venkatesan, S. y Moss, B. (1983). Structure and replication of vaccinia virus telomeres. Cold Spring Harbour Symps on Quant. Biol., pp. 723-729.
69. Gafford, L.G., Mitchell Jr., E.B. y Randall, C.C. (1978) Sedimentation characteristics and molecular weights of three poxvirus DNAs. Virology 89, 229-239.
70. Parr, R.P., Burnett, J.W. y Garon C.F. (1977) Structural characterization of the Molluscum Contagiosum virus genome Virology 81, 247-256.
71. Menna, A., Wittek, R., Bachmann, P.A., Mayr, A. y Wyler, R. (1979). Physical characterization of a Stomatitis papulosa virus genome: a cleavage map for the restriction endonuclease Hind III and Eco RI. Arch. Virol. 59, 145-156.
72. Maxam, A. y Gilbert, W. (1980). Sequencing end labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65, 499-599.
73. Moyer, R.W. y Graves, R.L. (1981) The mechanism of cytoplasmic orthopoxvirus DNA replication. Cell 27, 391-401.
74. Cavalier-Smith, T. (1974) Palindromic base sequences and replication of eukaryotic chromosomes Nature 250, 467-470.

75. Bateman, A.J.. (1975) Simplification of palindromic telomere theory. *Nature* 253, 379-380.
76. Tattersall, P. y Ward D. C. (1976). Rolling hairpin model for replication of parvovirus and linear chromosomal DNA. *Nature* 263, 106-109.
77. Esteban, M. y J.A. Holowez A.K. (1977). Replication of vaccinia DNA in mouse L cells I. in vivo DNA synthesis. *Virology* 78, 57-75.
78. Pogo, B.G.T. (1977). Elimination of naturally occurring crosslinks in vaccinia virus DNA after viral penetration into cells. *Proc. Natl. Acad. Sci.* 74, 1739-1742.
79. Pogo, B.G.T. (1980). Changes in parental vaccinia virus DNA after viral penetration into cells. *Virology* 101, 520.
80. Pogo, B.G.T. (1977). Replication of vaccinia DNA in mouse L. cells III Intracellular forms of viral DNA. *Virology* 82, 308.
81. Pogo B.G.T. y O'Shea M. (1978). The mode of replication of vaccinia virus DNA. *Virology* 84, 1.
82. Pogo B.G.T., O'Shea, M. y Freimuth P. (1981). Initiation and termination of vaccinia virus DNA replication. *Virology* 108, 241-248.
83. Forte M. A. y Fangman W. L. (1979). Yeast chromosomal DNA molecules have strands which are cross-linked at their termini. *Chromosoma* 72, 131-150.
84. Venkatesan S., Baroudy B. M. y Moss M. (1981). Distinctive nucleotide sequences adjacent to multiple initiation and termination sites of an early vaccinia virus gene. *Cell* 125, 805-813.

85. Weir J.P. y Moss B. (1983). Nucleotide sequence of the vaccinia virus thymidine kinase gene and the nature of spontaneous frameshift mutations. *J. Virol* 46, 530-537.
86. Hruby, D.E., Maki, R.A., Miller, D.B. y Ball, L.A. (1983). Fine structure analysis and nucleotide sequences of the vaccinia virus thymidine kinase gene. *Proc. Natl. Acad. Sci. USA.* 80, 3411-3415.
87. Venkatesan, S.A., Gerhowitz, y Moss B. (1982). Complete nucleotide sequences of adjacent early vaccinia virus genes located within the inverted terminal repetition. *J. Virol.* 44, 637-646.
88. Weir J.P. y Moss, B., (1984) Regulation of expression and nucleotide sequence of a late vaccinia virus gene. *J. Virol.* 51, 662-669.
89. Oda, K., y Joklik, W.K., (1967). Hybridization and sedimentation studies on "early" and "late" vaccinia messenger RNA. *J. Mol. Biol.* 27, 395-419.
90. Cooper, J.A. y Moss, B., (1979). In vitro translation of immediate early and late classes of RNA from vaccinia virus infected cells. *Virology* 96, 368-380.
91. Cooper, J.A., Wittek, R. y Moss, B. (1981). Extension of the transcriptional and translational map of the left end of the vaccinia virus genome to 21 kilobase pairs. *J. Virol.* 39, 733-745.
92. Mahr, A. y Robertd, B.E. (1984) Arrangement of late RNAs transcribed from a 7.1. kilobase EcoRI vaccinia virus DNA fragment. *J. Virol.* 49, 510-520.



93. Bajszar, G.R., Wittek, J.P., Weir y Moss, B. (1983). Vaccinia virus thymidine kinase and neighboring genes: mRNAs and polypeptides of wild-type virus and putative nonsense mutants. *J. Virol.* 45, 62-72.
94. Wittek, R., Cooper, J.A., Barbosa, E. y Moss, B. (1980). Expression of the vaccinia virus genome: analysis and mapping of mRNAs encoded within the inverted terminal repetition. *Cell* 21, 487-493.
95. Venkatesan, S. y Moss, B. (1981). In vitro transcription of the inverted terminal repetition of the vaccinia virus genome: correspondence of initiation and cap sites. *J. Virol* 37, 378-747.
96. Wittek, R., Hanggi, M. y Hiller, G. 1984 Mapping of a gene coding for a major late structural polypeptide on the vaccinia virus genome *J. Virol* 49, 371-378.
97. Plucienniczak, A., Schroeder, E., Zettlneissl, G. y Strebeck, R.E. (1985) Nucleotide sequence of a cluster of early and late genes in a conserved segment of the vaccinia virus genome. *Nucl. Ac. Res* 13, 985-998.
98. Kozak, M. (1984) Compilation of Sequences upstream from the translational start site in eukaryotic mRNAs. *Nucl. Ac. Res* 12, 857-872.
99. Puckett, C. y Moss, B. 1983 Selective transcription of vaccinia virus genes in template dependent soluble extracts of infected cells. *Cell* 35, 441-448.
100. Sanger, F., Nicklen, S. y Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors *Proc. Natl. Acad. Sci USA* 74, 5463-5467.

101. Biggin, M., Gibson, T.J. y Hong, G.F. (1983) Buffer gradient gels and  $^{35}\text{S}$  label as an aid to rapid DNA sequence determination. *Proc. Natl Acad. Sci. USA* 80, 3963-3966.
102. Beck, S. y Pohl, F.M. (1984). DNA sequencing with direct blotting electrophoresis *EMBO J.* 13, 2905-2909.
103. Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. y Peterson G.B. (1982) Nucleotide sequence of bacteriophage  $\lambda$  DNA *J. Mol. Biol.* 162, 729-773.
104. Baer, R., Bankier, A.T., Biggin, M.D., Deininger P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Stachwell, S.C., Seguin, C., Tufnell, P.S. y Barrell, B.G. (1984) DNA sequence and expresion of the B95-8 Epstein-Barr virus genome. *Nature* 310, 207-211.
105. Deininger, P.L. (1983) Approaches to rapid DNA sequence analysis. *Anal. Biochem.* 135, 247-263.
106. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. y Roe, B.A. (1980). Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol* 143, 151-178.
107. Winter, G. y Fields, S. (1980) Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. *Nucl. Ac. Res.* 8, 1965-1974.
108. Harley, J.L., Chen K.K. y Donelson, J.E. (1982). A mercury-thiol affinity system for rapid generation of overlapping labeled DNA fragment for DNA sequencing. *Nucl. Ac. Res.* 10, 4099-4025.
109. Barnes, W.M. y Bevan, M. (1983) Kilosequencing: an ordered strategy for rapid DNA sequence data acquisition. *Nucl. Ac. Res.* 11, 349-368.

110. Hong, G.F. (1982) A systematic DNA sequencing strategy  
J. Mol. Biol. 158, 539-549.
111. Frischauf, A.M., Garoff, H. y Lehrach, H. (1980). A sub-  
cloning strategy for DNA sequence analysis. Nucl. Ac. Res. 8,  
5541-5549.
112. Poncz, M., Solowiejczyk, D., Ballatine, M., Schwartz, E. y  
Surrey, S. (1982). Nonrandom DNA sequence analysis in bacterio-  
phage M13 by the dideoxi chain-termination method. Proc. Natl.  
Acad. Sci. USA 79, 4298-4302.
113. Messing, J., Gronenborn, B. Y Müller-Hill, B. (1977).  
Filamentous coliphage M13 as a cloning vehicle: insertion of a  
Hind II fragment of the lac regulatory region in the M13 repli-  
cative form in vitro. Proc. Natl. Acad. Sci USA 75, 3462-3646.
114. Messing, J., Crea, R. y Seeburg, PH. (1981). A system for  
shotgun DNA squencing Nucl. Ac. Res. 9, 309-321.
115. Vieira, J. y Messing, J. (1982) The pUC plasmids, an  
M13mp7-derived system for insertion mutagenesis and sequencing  
with synthetic universal primers. Gene 19, 259-268.
116. Yanisch-Perron, C., Vieira, J. y Messing, J. (1985)  
Improved M13 phage cloning vectors and host strains: nucleotide  
sequences oof the M13 mp18 and pUC19 vectors. Gene 33, 103-119.
117. Maniatis, T., Fritsch, E. F. y Sambrook, J.(1982). Mole-  
cular cloning. Ed. Cold Spring Harbor Laboratory.
118. Marinus, M. G. y Morris N. R. (1973). J. Bacteriol.  
114,1143-1150.
119. Birnboim, H.L. y Doly, J. (1979). A rapid alkaline ex-  
traction procedure for screening of recombinat plasmid DNA  
Nucl. Ac. Res. 7, 1513-1523.

120. Clewell, D.B. y Helsinki, D.R. (1970). Properties of a supercoiled DNA-protein relaxation complex and strand specificity of the relaxation event. *Biochemistry* 9, 4428-4440.
121. Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol Biol.* 98, 503-517.
122. Langridge, J., Langridge, P. y Bergquist, P.L. (1980) Extraction of nucleic acids from agarose gels *Anal. Biochem.* 103, 246-271.
123. Rigby, P., Dieckmann, M., Rhodes, C. y Berg, P. (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick-translation with DNA polymerase I. *J. Mol. Biol.* 113, 237-251.
124. Roop, D.R., Nordstrom, J.L., Tsai, S.Y., Tsai, M.J. y O'Malley, B.W. (1978). Transcription of structural and intervening sequences in the ovalbumin gene and identification of potential ovalbumin mRNA precursors. *Cell* 15, 671-685.
125. Roop, D.R., Tsai, M.J. y O'Malley, B.W. (1980). Definition of the 5' and 3' ends of transcripts of the ovalbumin gene. *Cell* 19, 63-68.
126. Wahl, G.M., Stern, M. y Stark, G.R. (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. *Proc. Natl. Acad. Sci. USA* 76, 3683-3687.
127. Mandel, M. y Higa, A. (1970). Calcium dependent bacteriophage DNA infection *J. Mol. Biol* 53, 159-162.
128. Morrison, D.A. (1979) Transformation and preservation of competent bacterial cells by freezing. *Meth Enzymol* 68, 326-331.

129. Grunstein, M. y Hogness, D.S. (1975) Colony hybridization: A method for the isolation of DNA clones that contain a specific gene. Proc. Natl. Acad. Sci. USA 72, 3961-3965.
130. Chou, P. Y. y Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their aminoacid sequence. Adv. Enzymol. Rel. Areas Mol. Biol. 47, 45-148.
131. Fitch, W. M. y Margoliash, E. (1967). Construction of phylogenetic trees. Science 155, 279-284.
132. Staden, R. (1984). Graphic methods to determine the function of nucleic acid sequences. Nucl. Ac. Res. 12, 521-538.
133. Hopp, T. P. y Woods, K. R. (1981). Prediction of protein antigenic determinants from aminoacid sequences. Proc. Natl. Acad. Sci. USA 78, 3824-3828.
134. Welling, G. W., Weijer, W. J., van der Zee, R. y Welling-Wester, S. (1985). Prediction of sequential antigenic regions in proteins. FEBS Lett. 188, 215-218.
135. Kyte, J. y Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.
136. Garnier, J. Osguthorpe, D. y Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120, 97-120.
137. Tinoco, I., Uhlenbeck, O. C. y Levine, M. D. (1971). Estimation of secondary structure in ribonucleic acids. Nature 230, 362-367.
138. Lipman, D. J. y Pearson, W. R. (1985). Rapid and sensitive protein similarity searches, Science 227, 1435-1441.

139. Lehninger, A. L. (1981). Bioquímica. Ed. Omega . Barcelona.
140. Gibbs A. y Fenner F. (1984). Methods for comparing sequence data such as restriction endonuclease maps or nucleotide sequences of viral nucleic acid molecules. J. Virol. Methods 9, 317-324.
141. McGeoch, D. J. (1985). On the predictive recognition of signal peptide sequences. Virus Res. 3, 271-286.
142. Perlman, D. y Halvorson, H. O. (1983). A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. 167, 391-409.
143. Sogo, J. M., Almendral, J. M., Talavera, A. y Viñuela, E. (1984). Terminal and inverted repetitions in ASFV DNA. Virology 133, 271-275.
144. Straus, S. E., Sebring, E. D. y Rose, J. A. (1976). Concatemers of alternating plus and minus strands are intermediates in adenovirus-associated virus DNA-synthesis. Proc. Natl. Acad. Sci. USA 73, 742-746.
145. Tattersall, P. y Ward, D. C. (1976). Rolling hairpin model for replication of parvovirus and linear chromosomal DNA. Nature 263, 106-109.
146. Goddard, J. M. y Cummings, D. J. (1977). Mitochondrial DNA replication in Paramecium aurelia. Cross-linking of the initial end. J. Mol. Biol. 190, 327-349.
147. Blackburn, E. H. y Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J. Mol. Biol. 120, 33-53.

148. Forte, M. A. y Fangman, W. L. (1979). Yeast chromosomal DNA molecules have strands which are cross-linked at their termini. *Cromosoma* 72, 131-150.
149. Berns, K. I., Hauswirth, W. W., Fife, K. H. y Lusby, E. (1979). Adeno-associated virus DNA replication. *Cold Spring Harbor Symp. Quant. Biol.* 43, 781.
150. Astell, C. R. Smith, M., Chou, M. B. y Ward, D. C. (1979). Structure of the 3' hairpin termini of four rodent parvovirus genomes: Nucleotide sequence homology at origin of DNA replication. *Cell* 17, 691-703.
151. Salzman, L. A. y Fabisch, P. (1979). Nucleotide sequence of the self priming of the single-stranded DNA extracted from the parvovirus Kilham rat virus. *J. Virol.* 30, 946-950.
152. Pritchard, A. E. y Cummings, D. J. (1981). Replication of linear mitochondrial DNA from *Paramecium*: Sequence and structure of the initiation end crosslink. *Proc. Natl. Acad. Sci. USA* 78, 7341-7345.
153. Blackburn, E. H. y Challoner, P. B. (1984). Identification of a telomeric DNA sequence in *Trypanosoma brucei*. *Cell* 36, 447-457.
154. Blackburn, E. H. (1984). Telomeres: do the ends justify the means? *Cell* 37, 7-8.
155. Szostak, J. W. y Blackburn, E. H. (1982). Cloning yeast telomeres on linear plasmid vectors. *Cell* 29, 245-255.
156. Shampay, J., Szostak, J. W. y Blackburn, E. H. (1984). DNA sequences of telomeres maintained in yeast. *Nature* 310, 154-157.
157. González, A., Talavera, A., Almendral, J. M. y Viñuela, E. Man. en prep.

158. González, A. com. pers.
159. Potter, S. S. (1982). DNA sequence of a foldback transposable element in *Drosophila*. *Nature* 297, 201-204.
160. Berg, O.G., Winter, R. y von Hippel, P. H. (1981). Diffusion driven mechanism of protein translocation on nucleic acids.1. Models and theory. *Biochemistry* 20, 6929-6948.
161. Viñuela, E. (1985). African swine fever virus. *Curr. Top. Microbiol. Immunol.* 116, 151-170.
162. Talavera, A. com. pers.
163. Santarén, J. F. y Viñuela, E. Man. en prep.
164. Grantham, R., Gautier, R., Corey, M., Jacobzone, M. y Mercier, R. (1981). Coding catalog usage and the genome hypothesis. *Nucl. Ac. Res.* 9, r43-r74.
165. Almendral, J. M., com. pers.
166. Blasco, R. (1985). Tesis doctoral. Universidad Autónoma. Madrid.
167. Carrascosa, A. L. , com. pers.
168. Moyer, R. W., Graves, L. R. y Rothe, C. T. (1980). The white pox mutants of rabbit poxvirus. III. Terminal DNA sequence duplication and transposition in RPV. *Cell* 22, 545-553.
169. Kant, J. A., Fornace, A. J. Jr, Saxe, D., Simon, M. I., McBride, O. W. y Crabtree, G. R. (1985). Evolution and organization of the fibrinogen locus on chromosome 4: gene duplication accompanied by transposition and inversion. *Proc, Natl. Acad. Sci. USA* 82, 2344-2348.
170. Lewin, B. (1983). *Genes*. Ed. John Wiley and sons. New York.




171. Forss, S. y Schaller, H. (1982). A tandem repeat gene in a picornavirus. Nucl. Ac. Res. 10, 6441-6450.

172. Dayhoff, M. O., Schwartz, R. M. y Orcutt, B. C.(1978). Atlas of protein sequence and structure. Vol. 5, supl. 3, 345-352.



~~APTO COM LAUNE~~

*E. J. Sam*



Auton Televa

F. G. Arnold

*Ernesto Escobar*

UNIVERSIDAD AUTONOMA DE MADRID



5408307705